PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04S 1/00, 5/00, 3/00

A2

(11) International Publication Number: WO 97/25834

(43) International Publication Date: 17 July 1997 (17.07.97)

(21) International Application Number: PCT/US97/00145

(22) International Filing Date: 3 January 1997 (03.01.97)

(30) Priority Data:

08/582,830	4 January 1996 (04.01.96)	US
08/726,518	7 October 1996 (07.10.96)	US
08/726,590	7 October 1996 (07.10.96)	US

(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

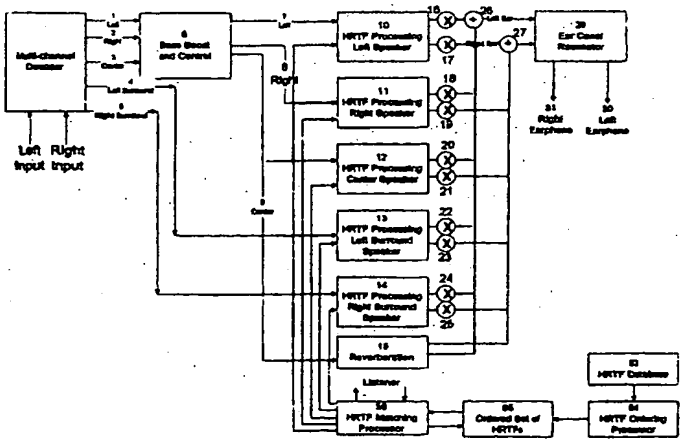
(71) Applicant *(for all designated States except US)*: VIRTUAL LISTENING SYSTEMS, INC. [US/US]; 4637 N.W. 6th Street, Gainesville, FL 32609 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: TUCKER, Timothy, J. [US/US]; 2215 N.W. 9th Place, Gainesville, FL 32609 (US). GREEN, David, M. [US/US]; 399 Federal Point Road, East Palatka, FL 32131 (US).

(74) Agents: PARKER, James, S. et al.; Saliwanchik, Lloyd & Saliwanchik, Suite A-1, 2421 N.W. 41st Street, Gainesville, FL 32606-6669 (US).

(54) Title: METHOD AND DEVICE FOR PROCESSING A MULTI-CHANNEL SIGNAL FOR USE WITH A HEADPHONE



(57) Abstract

A method and device processes multi-channel audio signals, each channel corresponding to a loudspeaker placed in a particular location in a room, in such a way as to create, over headphones, the sensation of multiple "phantom" loudspeakers placed throughout the room. Head Related Transfer Functions (HRTFs) are chosen according to the elevation and azimuth of each intended loudspeaker relative to the listener, each channel being filtered with an HRTF such that when combined into left and right channels and played over headphones, the listener senses that the sound is actually produced by phantom loudspeakers placed throughout the "virtual" room. A database collection of sets of HRTF coefficients from numerous individuals and subsequent matching of the best HRTF set to the individual listener provides the listener with listening sensations similar to that which the listener, as an individual, would experience when listening to multiple loudspeakers placed throughout the room. An appropriate transfer function applied to the right and left channel output allows the sensation of open-ear listening to be experienced through closed-ear headphones.

Best Available Copy

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

NSDOCID: <WO_____9725834A2_I_>

METHOD AND DEVICE FOR PROCESSING A MULTICHANNEL SIGNAL FOR USE WITH A HEADPHONE

Background of the Invention

Field of the Invention. The present invention relates to a method and device for processing a multi-channel audio signal for reproduction over headphones. In particular, the present invention relates to an apparatus and method for creating, over headphones, the sensation of multiple "phantom" loudspeakers in a user matched virtual listening environment.

Background Information. In an attempt to provide a more realistic or engulfing listening experience in the movie theater, several companies have developed multi-channel audio formats. Each audio channel of the multi-channel signal is routed to one of several loudspeakers distributed throughout the theater, providing movie-goers with the sensation that sounds are originating all around them. At least one of these formats, for example the Dolby Pro Logic® format, has been adapted for use in the home entertainment industry. The Dolby Pro Logic® format is now in wide use in home theater systems. As with the theater version, each audio channel of the multi-channel signal is routed to one of several loudspeakers placed around the room, providing home listeners with the sensation that sounds are originating all around them. As the home entertainment system market expands, other multi-channel systems will likely become available to home consumers.

When humans listen to sounds produced by loudspeakers, it is termed open-ear listening. Open-ear listening occurs when the ears are uncovered. It is the way we listen in everyday life. In an open-ear environment, the sonic information arriving at the ears provides cues about the location and distance of the sound source. Humans are able to localize a sound to the right or left based on differences in the arrival times and differences in the sound levels at the two ears. Other subtle differences in the spectrum of the sound at each ear drum provide cues about the sound source elevation and front/back location. These differences are related to the filtering effects of several body parts, most notably the head and the pinnae of the ears.

The process of listening while the outer ear surface of the ear is covered (e.g., with headphones) is termed closed-ear listening. Covering the ear changes the ear canal resonance characteristics. Due to the physical effects of wearing headphones, sound delivered through headphones lacks the subtle differences in time, level, and spectra caused by location, distance, and the filtering effects of the head and pinna experienced in open-ear listening. Thus, when headphones are used with multi-channel home entertainment systems, the advantages of listening via numerous

NSDOCID: <WO_____9725834A2_I_>

loudspeakers placed throughout the room are lost, the sound often appearing to be originating inside the listener's head.

There is a need for a system that can process multi-channel audio in such a way as to cause the listener to sense multiple "phantom" loudspeakers when listening over headphones. Such a system should process each channel such that the effects of loudspeaker location and distance intended to be created by each channel signal, as well as the filtering effects of the listener's head and pinnae are preserved or simulated accurately for that individual listener.

Accordingly, an object of the present invention is to provide a method for processing the multi-channel output typically produced by home entertainment or like systems such that when presented over headphones, the listener is able to select a best match set of head related transfer functions from a database of measured head related transfer functions to filter the channels such that the listener experiences the sensation of multiple "phantom" loudspeakers placed throughout the room.

Another object of the present invention is to provide an apparatus for processing the multi-channel output typically produced by home entertainment or like systems such that when presented over headphones, the listener experiences listening sensations most like that which the listener, as an individual, would experience when listening to multiple loudspeakers placed throughout the room.

Another object of the present invention is to provide an apparatus for processing the multi-channel output typically produced by home entertainment or like systems such that when presented over headphones, the listener experiences sensations typical of open-ear (unobstructed) listening.

Another object of the present invention is to provide an apparatus and method for measuring the acoustic filtering action produced by the head and pinnae of the human ears so as to produce a useful database of head related transfer functions.

Another object of the present invention is to create a database of HRTFs representative of the general listening public by measuring and recording a large enough set of such HRTFs such that any given individual is likely to be able to select a set of HRTFs from the database so that when used to process an audio signal the user perceives the corresponding sounds to be localized in the proper spatial positions.

Another object of the present invention is to provide a means of determining the "best-match" of an individual listener to one of the HRTF sets of the representative database such that the individual listener can be matched as closely as possible to an already measured set of HRTFs stored in a database, such that once properly matched, the individual will experience the correct "phantom" locations of the sources of the listening system.

Another object of the present invention is to provide a wired or wireless transmission system for dimensionalized listening of sound over headphones.

Other objects of the invention will become clear from a review of the complete disclosure.

Summary of the Invention

According to the present invention, multiple channels of an audio signal are processed through the application of filtering using a head related transfer function (HRTF) or a plurality of HRTFs, selected by a user, such that when reduced to two channels, left and right, each channel contains information that enables the listener to sense the location of multiple phantom loudspeakers when listening over headphones.

Also according to the present invention, multiple channels of an audio signal are processed through the application of filtering using HRTFs chosen from a large database such that when listening through headphones, the listener experiences a sensation that most closely matches the sensation the listener, as an individual, would experience when listening to multiple loudspeakers.

In another exemplary embodiment of the present invention, the right and left channels are filtered in order to simulate the effects of open-ear listening.

In another exemplary embodiment of the present invention, a complete set of HRTFs for an individual is measured and recorded, such that the measured HRTFs are an accurate reflection of the filtering effects of that individual's head and pinnae, and in which the measurement takes on the order of a few minutes. For each individual, several hundred HRTFs are measured such that an HRTF is specified for each location in space about the listener with an accuracy of approximately 10° in both the vertical and horizontal dimensions.

In a further embodiment of this invention, the HRTFs of a sufficient number of individuals are measured and stored to create a database such that a given individual is able to select a set of HRTFs from the database such that when audio signals are processed with the selected set of HRTFs, the user perceives the corresponding sounds to be localized in the proper spatial positions.

In a further embodiment, the database of HRTFs comprises a representative set of HRTF sets.

In another exemplary embodiment of the present invention, an individual is matched to a "best-match" set of HRTFs selected from a database of sets of HRTFs measured from a representative sample of the general listening population, where the individual listener participates in the matching of the set of HRTFs by comparing the perception created by different HRTF sets and selecting the HRTF set providing the best spatial perception.

NSDOCID: <WO_____9725834A2_I_>

In another exemplary embodiment of the present invention, a database of HRTF sets, measured from a representative sample of the listening population, is established, such that an individual can select a "best-match" set of HRTFs from the database.

In a further embodiment a best match set of HRTFs is selected from the database of HRTFs and is used to process signals for wired or wireless transmission to a listener wearing headphones.

Brief Description of the Drawings

Figure 1 is a representation of sound waves received at both ears of a listener sitting in a room with a typical multi-channel loudspeaker configuration.

Figure 2 is a representation of the listening sensation experienced through headphones according to an exemplary embodiment of the present invention.

Figure 3a shows the sound source locations used to measure a set of head related transfer functions (HRTFs) obtained at multiple elevations and azimuths surrounding a listener.

Figure 3b is a graph representing the HRTF for 0 degrees elevation and 30 degrees azimuth for three different individuals.

Figure 4 is a schematic in block diagram form of a typical multi-channel headphone processing system according to an exemplary embodiment of the present invention.

Figure 5 is a schematic in block diagram form of a bass boost circuit according to an exemplary embodiment of the present invention.

Figure 6A is a schematic in block diagram form of HRTF filtering as applied to a single channel according to an exemplary embodiment of the present invention.

Figure 6B is a schematic in block diagram form of the process of HRTF matching based on an ordered set of HRTFs according to the present invention.

Figure 7 is a representation of a typical digital signal transmission system comprising a transmitting station, a connecting medium called a channel and a receiving station.

Figure 8A is a block diagram of a novel radio-frequency transmission system for use in a wireless embodiment of this invention.

Figure 8B is a representation of an adaptive filter for removing the DC component of a digital signal.

Figure 9A shows a computer simulated input gaussian noise source with a variance of 2.5 mV and a mean of 0.5 V.

Figure 9B shows the tracking constant, C[k], during a computer simulation of the removal of the DC component of an input gaussian noise source by an adaptive filter.

Figure 9C shows the output of an adaptive filter where the input is a gaussian noise source.

Figures 9D and **9E** show the magnitude frequency response of the input gaussian noise waveform and DC shifted output.

Figure 9F is a schematic of a state machine.

Figure 9G is a timing diagram of various clock outputs for decoding signals encoded according to one embodiment of this invention.

Figure 10 depicts an HRTF matching process according to the present invention.

Figure 11 shows an impulse response wave form recorded from one individual at one spatial location for one ear.

Figure 12 illustrates critical band filtering according to the present invention.

Figure 13 illustrates an exemplary subject filtered HRTF matrix according to the present invention.

Figure 14 illustrates a hypothetical hierarchical agglomerative clustering procedure in two dimensions according to the present invention.

3NSDOCID: <WO 9725834A2 I >

Figure 15 illustrates a hypothetical hierarchical agglomerative clustering procedure according to an exemplary embodiment of the present invention.

Figure 16 is a schematic in block diagram form of a typical reverberation processor constructed of parallel lowpass comb filters.

Figure 17 is a schematic in block diagram form of a typical lowpass comb filter.

Figure 18a is a schematic of a preferred embodiment of an HRTF measurement means.

Figure 18b further illustrates a preferred embodiment of an HRTF measurement means.

Figure 19 is a schematic representation of the HRTF measurement control system.

Figure 20 is a schematic representation of the HRTF measurement control system software flow chart.

Figure 21A is a schematic representation of a front view of a sound room in which HRTFs may be measured to produce the database of HRTFs of this invention.

Figure 21B is a schematic representation of a top view of the sound room.

Figure 21C shows the detail of the cross section of the wall of the sound room.

Figure 22A shows the probability that the RMS distance, between any individual's HRTF and the nearest HRTF already in the database, is less than a certain RMS distance (dB), as a function of the number of HRTF sets in the database.

Figure 22B shows the cumulative density function of the distance between each of 150 HRTFs and the mean HRTF.

Figure 22C shows the change in average mean as a function of subsample group size.

Figure 22D shows the change in average standard deviation as a function of subsample group size.

Figure 22E shows the mean minimum distance between any HRTF set of the 150 HRTF sets and one of the stored HRTF sets as a function of the number of stored HRTF sets.

Figures 23A, B, C are block diagrams of a circuit according to this invention for processing signals using a best match set of HRTFs selected by a user from the database of this invention.

Figure 24 is a detail of an early reflection processing circuit **612** according to Figure 23.

Figure 25 is a detail of an HRTF processing circuit **663** according to Figure 23 comprising finite impulse response filters that implement HRTFs selected from the database of this invention.

Figure 26 is a detail of a reverberation circuit **671** according to Figure 23.

Figure 27 is a detail of a bass boost processing circuit **670** according to Figure 23.

Figures 28A, B, C are a schematic representation of the HRTF selection and matching performed by a user to arrive at a best match set of HRTFs which is then used for processing of audio signals according to Figures 25 and 23.

Figure 29A, B is an alternate embodiment to that disclosed in Figures 28A, B, and C.

Detailed Description of the Invention

The method and device according to the present invention processes audio signals, including multi-channel audio signals having a plurality of channels, each corresponding to a loudspeaker placed in a particular location in a room, in such a way as to create, over headphone, the sensation of multiple "phantom" loudspeakers placed throughout the room. The present invention utilizes Head Related Transfer Functions (HRTFs) that are chosen according to the elevation and azimuth of each intended loudspeaker relative to the listener, each channel being filtered by a set of HRTFs such that when combined into left and right channels and played over headphones, the listener

NSDOCID: <WO_____9725834A2_I_>

senses that the sound is actually produced by phantom loudspeakers placed throughout the "virtual" room.

The filtering of the present invention utilizes a database collection of sets of HRTFs measured from numerous individuals and subsequent matching of the best HRTF set to an individual listener, thus providing the listener with listening sensations similar to that which the listener, as an individual, would experience when listening to multiple loudspeakers placed throughout the room. Additionally, the present invention utilizes an appropriate transfer function applied to the right and left channel output so that the sensation of open-ear listening may be experienced through closed-ear headphones.

In generating the database collection of sets of HRTFs, the present invention also provides a measurement device and method for measuring and recording complete sets of HRTFs of subjects from a representative sample of the listening population, such that the measured HRTFs are an accurate reflection of the filtering effects of the head and pinnae of each of the subjects measured. For each individual, as many as 360 HRTFs for each ear may be measured, with each HRTF depending on the position or location of the sound source with respect to the listener. These measured HRTF sets are stored in a database, such that the database provides HRTF sets from which any individual can select a set of HRTFs such that when audio signals are processed with the selected set of HRTFs, the user perceives the corresponding sounds to be localized in the proper spatial positions, to thereby achieve optimized 3D virtual audio effects when using headphones.

Figure 1 depicts the path of sound waves received at both ears of a listener according to a typical embodiment of a home entertainment system. The multi-channel audio signal is decoded into multiple channels, i.e., a two-channel encoded signal is decoded into a multi-channel signal in accordance with, for example, the Dolby Pro Logic® format. Each channel of the multi-channel signal is then played, for example, through its associated loudspeaker, e.g., one of five loudspeakers: left; right; center; left surround; and right surround. The effect is the sensation that sound is originating all around the listener.

Figure 2 depicts the listening experience created by an exemplary embodiment of the present invention. As described in detail with respect to Figure 4, the present invention processes each channel of a multi-channel signal using a set of HRTFs appropriate for the distance and location of each phantom loudspeaker (e.g., the intended loudspeaker for each channel) relative to the listener's left and right ears. All resulting left ear channels are summed, and all resulting right ear channels are summed producing two channels, left and right. Each channel is then preferably filtered using a transfer function that introduces the effects of open-ear listening. When the two channel output

is presented via headphones, the listener senses that the sound is originating from five phantom loudspeakers placed throughout the room, as indicated in Figure 2.

The manner in which the ears and head filter sound may be described by a Head Related Transfer Function (HRTF). An HRTF is a transfer function obtained from one individual for one ear for a specific sound source location. An HRTF is described by multiple coefficients that characterize how sound produced at a particular spatial position should be filtered to simulate the filtering effects of the head and outer ear of a particular individual. HRTFs are typically measured at various elevations and azimuths. Typical HRTF measurement locations are illustrated in Figure 3A.

In Figure 3A, the horizontal plane located at the center of the listener's head **100** represents 0.0° elevation. The vertical plane extending forward from the center of the head **100** represents 0.0° azimuth. HRTF locations are defined by a pair of elevation and azimuth coordinates and are represented by a small sphere **110**. In one embodiment of this invention, HRTFs are measured in 10 degree intervals for the azimuth and 10 degree intervals for the elevation from 30 degrees below the horizon to 60 degrees above the horizon. Associated with each sphere **110** is a set of HRTF coefficients that represent the transfer function for that sound source location. Each sphere **110** is actually associated with two HRTFs, one for each ear.

Because no two humans have identical heads and pinnae, no two humans have HRTFs which are exactly alike. This fact is demonstrated in Figure 3B which shows a graph representing the HRTF for 0 degrees elevation and 30 degrees azimuth for three different individuals. As can be seen, each of these individuals has quite different HRTFs. Therefore, for each individual, it is critical to use a set of HRTFs for filtering audio signals such that when the audio signals are filtered, the user perceives the corresponding sounds to be localized in the proper positions, in order to optimally create the sensation that the particular signal originates from the location which is intended by the HRTF processing. There have been some efforts to use a "universal" set of HRTFs, wherein every user is presented with the same set of HRTFs, having some average characteristics. However, as one can see from Figure 3B, a "universal" set of HRTFs would give very different sensations to each of the three individuals depicted. For instance, if an individual's HRTF had a peak (or valley) at a frequency f, while the universal HRTF had a contradictory valley (or peak) at the same frequency f, the individual would interpret the directional cues of the signal incorrectly. These inaccurate or poorly matched HRTFs degrade the overall 3D perception of the individual, the amount of degradation depending on the individual. This was experimentally demonstrated by Wightman and Kistler (1993).

NSDOCID: <WO_____9725834A2_I_>

In order to improve performance beyond the use of a single or "universal" HRTF, and to overcome the impracticalities of measuring an individual set of HRTFs for each individual, the present invention provides a database of HRTFs collected from a measured group of the general population. For example, the HRTFs are collected from numerous individuals of both sexes with varying physical characteristics. The present invention then employs a unique process whereby the sets of HRTFs obtained from all individuals are organized into an ordered fashion and stored in a read only memory (ROM) or other storage device. An HRTF matching processor enables each user to select, from the sets of HRTFs stored in the ROM, a set of HRTFs such that when audio signals are processed with the selected set of HRTFs, the user perceives the corresponding sounds to be localized in the proper spatial positions.

An exemplary embodiment of the present invention is illustrated in Figure 4. After the multi-channel signal has been decoded into its constituent channels, for example channels 1, 2, 3, 4 and 5 in the Dolby Pro Logic® format, selected channels are processed via an optional bass boost circuit 6. For example, channels 1, 2 and 3 are processed by the bass boost circuit 6. Output channels 7, 8 and 9 from the bass boost circuit 6, as well as channels 4 and 5, are then each electronically processed to create the sensation of a phantom loudspeaker for each channel.

Processing of each channel is accomplished through digital filtering using sets of HRTF coefficients, for example via HRTF processing circuits 10, 11, 12, 13 and 14. The HRTF processing circuits can include, for example, a suitably programmed digital signal processor. A best match between the listener and a set of HRTFs is selected via the HRTF matching processor 59. Based on the best match set of HRTFs, a preferred pair of HRTFs, one for each ear, is selected for each channel as a function of the intended loudspeaker position of each channel of the multi-channel signal. In an exemplary embodiment of the present invention, the best match set of HRTFs are selected from an ordered set of HRTFs stored in ROM 65 via the HRTF matching processor 59 and routed to the appropriate HRTF processor 10, 11, 12, 13 and 14.

Prior to the listener selecting a best match set of HRTFs, sets of HRTFs stored in the HRTF database 63 are processed by an HRTF ordering processor 64 such that they may be stored in ROM 65 in an ordered sequence to optimize the matching process via HRTF matching processor 59. Once the optimal pair of HRTFs for each channel have been selected by the listener, separate HRTFs are applied for the right and left ears, converting each input channel to dual channel output.

Each channel of the dual channel output from, for example, the HRTF processing circuit 10 is multiplied by a scaling factor as shown, for example, at nodes 16 and 17. This scaling factor reflects signal attenuation as a function of the distance between the phantom loudspeaker and the listener's ear. All right ear channels are summed at node 26. All left ear channels are summed at

NSDOCID: <WO_____9725834A2_I_>

node 27. The output of nodes 26 and 27 results in two channels, left and right respectively, each of which contains signal information necessary to provide the sensation of left, right, center, and rear loudspeakers intended to be created by each channel of the multi-channel signal, but now configured to be presented over conventional two transducer headphones.

Additionally, parallel reverberation processing may optionally be performed on one or more channels by reverberation circuit 15. In a free-field, the sound signal that reaches the ear includes information transmitted directly from each sound source as well as information reflected off of surfaces such as walls and ceilings. Sound information that is reflected off of surfaces is delayed in its arrival at the ear relative to sound that travels directly to the ear. In order to simulate surface reflection, at least one channel of the multi-channel signal would be routed to the reverberation circuit 15, as shown in Figure 4.

In an exemplary embodiment of the present invention, one or more channels are routed through the reverberation circuit 15. The circuit 15 includes, for example, numerous lowpass comb filters in parallel configuration. This is illustrated in Figure 16. The input channel is routed to lowpass comb filters 140, 141, 142, 143, 144 and 145. Each of these filters is designed, as is known in the art, to introduce the delays associated with reflection off of room surfaces. The output of the lowpass comb filters is summed at node 146 and passed through an allpass filter 147. The output of the allpass filter is separated into two channels, left and right. A gain, g, is applied to the left channel at node 147. An inverse gain, -g, is applied to the right channel at node 148. The gain g allows the relative proportions of direct and reverberated sounds to be adjusted.

Figure 17 illustrates an exemplary embodiment of a lowpass comb filter 140. The input to the comb filter is summed with filtered output from the comb filter at node 150. The summed signal is routed through the comb filter 151 where it is delayed D samples. The output of the comb filter is routed to node 146, shown in Figure 16, and also summed with feedback from the lowpass filter 153 loop at node 152. The summed signal is then input to the lowpass filter 153. The output of the lowpass filter 153 is then routed back through both the comb filter and the lowpass filter, with gains applied of g_1 and g_2 at nodes 154 and 155, respectively.

The effects of open-ear (non-obstructed) resonation are optionally added at circuit 29 in Figure 4. The ear canal resonator according to the present invention is designed to simulate open-ear listening via headphones by introducing the resonances and anti-resonances that are characteristic of open-ear listening. It is generally known in the psychoacoustic art that open-ear listening introduces certain resonances and anti-resonances into the incoming acoustic signal due to the filtering effects of the outer ear. The characteristics of these resonances and anti-resonances are also generally known and may be used to construct a generally known transfer function, referred to as the

INSDOCID: <WO_____9725834A2_I_>

open ear, transfer function, that, when convolved with a digital signal, introduces these resonances and anti-resonances into the digital signal.

Open-ear resonation circuit 29 compensates for the effects introduced by obstruction of the outer ear via, for example, headphones. The open ear transfer function is convolved with each channel, left and right, using, for example, a digital signal processor. The output of the open-ear resonation circuit 29 is two audio channels 30, 31 that when delivered through headphones, simulate the listener's multi-loudspeaker listening experience by creating the sensation of phantom loudspeakers throughout the simulated room in accordance with loudspeaker layout provided by format of the multi-channel signal. Thus, the ear resonation circuit according to the present invention allows for use with any headphone, thereby eliminating a need for uniquely designed headphones.

Sound delivered to the ear via headphones is typically reduced in amplitude in the lower frequencies. Low frequency energy may be increased, however, through the use of a bass boost system. An exemplary embodiment of a bass boost circuit 6 is illustrated in Figure 5. Output from selected channels of the multi-channel system is routed to the bass boost circuit 6. Low frequency signal information is extracted by performing a low-pass filter at, for example, 100 Hz on one or more channels, via low pass filter 34. Once the low frequency signal information is obtained, it is multiplied by predetermined factor 35, for example k, and added to all channels via summing circuits 38, 39 and 40, thereby boosting the low frequency energy present in each channel.

To create the sensation of multiple phantom loudspeakers over headphones, the HRTF coefficients associated with the location of each phantom loudspeaker relative to the listener must be convolved with each channel. This convolution is accomplished using a digital signal processor and may be done in either the time or frequency domains with filter order ranging from 16 to 32 taps. Because HRTFs differ for right and left ears, the single channel input to each HRTF processing circuit 10, 11, 12, 13 and 14 is processed in parallel by two separate HRTFs, one for the right ear and one for the left ear. The result is a dual channel (e.g., right and left ear) output. This process is illustrated in Figure 6A.

Figure 6A illustrates the interaction of HRTF matching processor 59 with, for example, the HRTF processing circuit 10. Using the digital signal processor of HRTF processing circuit 10, the signal for each channel of the multi-channel signal is convolved with two different HRTFs. For example, Figure 6A shows the left channel signal 7 being applied to the left and right HRTF processing circuits 43, 44 of the HRTF processing circuit 10. One set of HRTF coefficients corresponding to the spatial location of the phantom loudspeaker relative to the left ear is applied to signal 7 via left ear HRTF processing circuit 43, the other set of HRTF coefficients corresponding

to the spatial location of the phantom loudspeaker relative to the right ear and being applied to signal 7 via the right ear HRTF processing circuit 44.

The HRTFs applied by HRTF processing circuits 43, 44 are selected from the set of HRTFs that best matches the listener via the HRTF matching processor 59. The output of each circuit 43, 44 is multiplied by a scaling factor via, for example, nodes 16 and 17, also as shown in Figure 4. This scaling factor is used to apply signal attenuation that corresponds to that which would be achieved in a free field environment. The value of the scaling factor is inversely related to the distance between the phantom loudspeaker and the listener's ear. As shown in Figure 4, the right ear output is summed for each phantom loudspeaker via node 26, and left ear output is summed for each phantom loudspeaker via node 27.

Once the left and right channel signals are processed and contain signal information necessary to provide the intended multi-channel sensation, the signal can be transmitted to conventional two transducer headphones. These signals can be transmitted by wire or wirelessly, for example, by a radio frequency (RF) transmission system. Examples of wireless transmission systems are exemplified in Examples 2, 3, and 4.

A central feature of this invention is to provide a sufficiently diverse and comprehensive set of HRTFs so that the user can select from that set one HRTF set which will produce the perception of sound located in the proper spatial position. This selection process is accomplished herein by: (1) collecting a comprehensive database of HRTFs; (2) ordering the database so that a representative subset of the entire collection of HRTFs can be obtained and stored in the device; and (3) providing a means for a user to select from the representative subset.

As described earlier, a single HRTF (see Figure 3B) is the spectrum obtained by presenting sound from a single location 110 (see Figure 3A). A listener's HRTF (head related transfer function) refers to the set of HRTFs obtained from the multiple locations described, for example, in Figure 3A. For any source location, two HRTFs are measured, one for the listener's left ear and one for the right ear. Thus, if L locations are measured, the set of 2*L spectra represent the HRTF set for a single listener. If S subjects are measured, an entire data base consisting of S*L*2 spectra is generated. In one embodiment, 360 locations (L=360) were measured and HRTFs on over 150 subjects were collected. Thus, the total data base consists of more than 108,000 spectra. These, or representative spectra are chosen (see below), and are stored in a database 63 (see Figures 4 and 6B).

For collecting these spectra a special robot arm was constructed. Prior measurement devices involved the use of multiple, e.g., 12, loudspeakers located on a circular hoop. Each of the multiple loudspeakers were used to create a signal used to measure the head-ear filter characteristics. In using these prior measurement devices, signals from each of the multiple loudspeakers were projected from

NSDOCID: <WO 9725834A2 >

a different location to allow measurements of HRTFs for different elevations and azimuths. However, the use of multiple loudspeakers poses a problem. To avoid contamination of the measured HRTF, the different loudspeakers need to have equal output spectra. Unfortunately, it is only possible to equate such spectra to within about 0.5 dB.

Advantageously, in the present invention, an improved measurement method is provided by utilizing a single loudspeaker located at the end of a robot arm. The single loudspeaker is used for all HRTF measurements, thereby eliminating the problem of unequal output spectra of different loudspeakers. The single loudspeaker is precisely positioned by a computer-controlled robot arm in each of the locations where an HRTF is to be measured. The present HRTF measurement device can measure and record a complete set of 360 HRTFs for each ear, for an individual, in approximately 10 to 15 minutes, as compared to one-to-four hours for prior measurement techniques. Because the listener should remain stationary during the entire measurement process, the speeding-up of the measurement process can, itself, contribute to the accuracy of the measurements.

Provided in Figure 18A is a schematic of a preferred embodiment of an HRTF measurement means according to this invention. At **200** there is provided a speaker, preferably a 4 Ohm, 40 watt speaker, for example, produced by Pioneer. At **201**, there is provided a lower arm, with dimensions approximately 1" wide, about 2" high and about 29" long. At **202**, there is provided an elbow AC servo motor, preferably capable of high rotational speeds and torques (e.g. about 20,000 rpm, and about 200 oz.-in.), and an absolute encode (e.g. about 500 count/rev.). Affixed to the elbow AC servo motor, there is provided an elbow planetary gearbox **203**, preferably with a ratio of about 100:1 and a torque capability of about 275 in.- lb. An upper arm **212** is connected to the lower arm **201** through the elbow AC servo motor **202**. At the upper end of the upper arm **212**, there is provided a shoulder spur gear pair **204**, preferably having a ratio of about 11.1111:1. Maintaining the shoulder spur gear in appropriate linkage with the upper arm **212** is a mounting bracket with bearings **205**. The mounting bracket **205** is suspended from a rotation shaft **206** having a diameter of about 1-1/4". A rotation spur gear pair **207** is provided with a ratio of about 12.8:1, to rotate the rotation shaft **206**. A rotation planetary gearbox **208**, having a ratio of about 100:1 and a torque capability of about 275 in. - lb., drives the rotation spur gear pair **207**. A rotation servo motor and associated absolute encoder **209** having a speed of about 20,000 rpm, a torque of about 200 oz. - in., with the encoder being amenable to 500 count/rev., are provided to actuate the rotation planetary gearbox **208**. A shoulder planetary gearbox **210**, having a ratio of about 100:1 and a torque output of about 275 lb. -in., is actuated by an associated shoulder servo motor **211** having a speed of about 20,000 rpm and a torque output of about 200 oz. - in. and an absolute encoder capable of about 500

count/rev., are linked to the shoulder spur gear **204** through a drive shaft **214**. A wrist gearmotor **213** having a speed of about 50 rpm and a torque of about 178 oz. - in. with an associated analog encoder are provided to position to the speaker **200**.

In Figure 18B, there is provided a detail of the upper arm **212**, the elbow planetary gearbox **203**, the elbow AC servo motor and absolute encode **202**, the mounting bracket with bearings **205**, the rotation shaft **206**, the shoulder planetary gearbox **210**, the shoulder servo motor and absolute encoder **211** and the drive shaft **214**.

In Figure 19, there is provided a schematic representation of the HRTF measurement control system. This includes a central control computer **300** which, in a first loop, controls a servo controller **301** which drives a plurality of servo amps **302a-c**, which in turn drive a plurality of linked encoder, servo motor and gearboxes **303a-c**. Encoder/servo motor/gearbox **303a** drives rotation, while **303b** drives the shoulder, and **303c** drives the arm (see Figure 18). In a second loop, the central control computer **300** controls data acquisition, signal presentation and speaker control via a feedback loop comprising: an encoder/gear/motor assembly **304** for positioning the speaker **305**; an A/D converter **306**, a D/A converter **307**, and an attenuator **308**. The feedback loop links through an amplifier **309** to the speaker **305** and to a microphone pre-amplifier **310** and the left and right microphones **311a** and **311b**. It will be appreciated that the above described hardware, and in particular the specifics of the various motor and gear power, rotation rates and ratios are all subject to modifications without adversely affecting the general principal of rapid, automated HRTF data acquisition with improved accuracy.

The above described hardware may be controlled by software which controls the positioning of the speaker. A preferred embodiment of such software is schematically represented in Figure 20. As can be seen, the software controls system startup at **400**, system initialization **401**, and display of a main menu **402**. Subroutines **403-408** are provided which allow for loading of data **403**, speaker calibration **404**, headphone measurement **405**, performance of an HRTF test run **406**, performance of a full HRTF measurement run **407**, and termination of the program **408**. A schematic of a full HRTF measurement run **407** is shown in steps **407a-407q**, all of which are initiated by selection of element **407** at the main menu. At **407a** the full HRTF measurement run is initiated, following which the measured subject is identified **407b**, the robot arm is calibrated **407c**, via a feedback loop **407d** which repeats arm calibration until a calibration "OK" signal **407e** is received. The robot arm is set to a zero starting position **407f**, and the measurement routine is begun **407g**. This includes movement of the robot arm and speaker **407 h** about the subject whose HRTF sets are being measured. The acquired data is played/recorded **407i** and the HRTF azimuth and elevation is displayed **407j** on a monitor. A continuous interrupt query **407k** is sent and as long as

3NSDOCID: <WO_____9725834A2_I_>

no interrupt signal is received, the measurement process is looped **407l** back to measurement step **407g**. If an interrupt signal is received, the system resets **407p** to the main menu, **407q**. If the measurement routine is continued without interruption, a complete set of HRTFs are measured until the natural termination of the measurement routine is reached **407m**. A pause **407n** is included in the routine to allow the system to store **407o** the acquired HRTFs, after which the system resets to the main menu **407q**.

The headphone measurement **405** comprises steps **405a-405h**, which are initiated by selecting this option at the main menu: at **405a**, the routine is initiated, following which sounds are played through the headphone and displayed **405b**. A pause **405c** is included in the routine to allow time for data retrieval and initiation of a subroutine **405d**. If a particular headphone subroutine is not to be initiated **405e** the system resets to the main menu. However, if a particular headphone subroutine is to be initiated, a particular headphone identity is entered **405f** and the data acquired for that headphone is stored **405g** following which the system resets to the main menu **405h**.

Optimally, the HRTF measurements are made in an appropriately constructed sound room. In a preferred embodiment of this invention, the measurements are made in a room such as that schematically depicted in Figures 21A, 21B, and 21C. This room, shown in a front view in Figure 21A, provides an exhaust fan **500** and an air outlet channel **510**. A latched door **520** is provided, preferably with latches on both the inside and outside. A fresh air fan **530** is provided for replenishment of fresh air from the outside of the room through an air inlet channel **540**. In Figure 21B, a schematic of a top view of the sound room is provided, including a representation of the subject seat **550**, a monitoring camera **560**, a pair of laser pointers **570**, and sound absorbent walls **580**. In Figure 21C a detail of the wall cross section is provided, showing a double wall structure in which there is provided two layers of dry wall **581** between which there is placed a damping material **582**, preferably selected from foam rubber, polyurethane or like sound insulating material.

A further improvement in the present HRTF measurement device and method is the location of the transducer employed to record the sound signal used in calculating the HRTF. Prior measurement techniques attempted to measure the sound as close to the eardrum as possible, by placing a narrow tube deep into the outer ear canal to measure the HRTF just at the eardrum. However, through physical considerations of the nature of sound transmission and the fact that the ear canal is small, we conclude that only a plane wave travels in the ear canal below frequencies of about 23,000 to 26,000 hertz. Since only plane waves travel in the ear canal at these frequencies, we expect that there is no directional information derived from the effect of the ear canal on the incoming sound. Since no directional information is derived from propagation of the sound down the ear canal, in the present HRTF measurement device and method, the transducer may be placed

at the entrance of the outer ear canal, instead of deep into the outer ear canal near the eardrum. In addition to being less uncomfortable for the individual "wearing" the transducer, the external location of the transducer provides a much higher S/N ratio than previous locations for the transducer. This higher S/N ratio provides a more accurate HRTF, especially in the "valleys" of the HRTF where the greatest attenuation of the incoming impulse signal exists.

The database of measured HRTF's is ordered by comparing the spectra recorded from different individuals. This is accomplished by transforming or pre-processing the raw data to represent the perceptual features of the raw spectra more accurately. The raw HRTFs are measured as the impulse response to a digital signal propagated by a loudspeaker at a given location. The signal so generated is carefully measured in the free-field (in the listener's absence) to correct for imperfections in the spectrum of the loudspeaker. The measured impulse response is then converted to the frequency domain using a fast Fourier transform (FFT) according to methods well known in the art. This frequency domain representation is further processed by implementing critical-band filtering and converting the data from a linear frequency scale to a logarithmic scale. Critical-band filtering reflects the fact that the first stage of the auditory system contains bandpass filters whose bandwidth is a constant fraction of the center frequency of the filter. The critical band filters resemble 1/6 octave bandpass filters. In addition, the distance along the auditory display is roughly proportional to the logarithm of sound frequency. Therefore, a logarithmic, rather than a linear, frequency scale is imposed on the representation.

In an exemplary embodiment, a gammatone filter is used to perform critical band filtering. The magnitude of the frequency response is represented by the function;

$$g(f) = 1/(1+[(f-fc)^2/b^2])^2$$

where f is frequency, fc is the center frequency for the critical band and b is 1.019 ERB. ERB varies as a function of frequency such that ERB = 24.7[4.37(fc/1000)+1]. For each critical band filter, the magnitude of the frequency response is calculated for each frequency, f, and is multiplied by the magnitude of the HRTF at the same frequency, f. For each critical band filter, the results of this calculation at all frequencies are squared and summed. The square root is then taken. This results in one value representing the magnitude of the internal HRTF for each critical band filter.

The hearing system is sensitive to a fixed fractional change in signal magnitude, which is known in the field as "Weber's Law." Thus, if stimulus magnitude is represented on a logarithmic scale, such as decibels, the ear is sensitive to a fixed number of decibels. In sum, the internal spectrum is represented by the level of the stimulus in decibels at about 12-18 frequencies per octave in the range between 3 and 18 kHz. Outside this frequency range (3 to 18 kHz) the human auditory system gains little or no directional or localization information based on the shape of the stimulus

NSDOCID: <WO_____9725834A2 I >

spectrum. In fact, few listeners but the very young can hear sounds above 18,000 Hz. At the lower frequencies, the spectrum of the signal is essentially the same for any azimuth or elevation. At the lower frequencies, however, especially below 4 kHz, differences in time of arrival at the two ears (interaural time cues) are important to indicate differences in the azimuthal position of the source.

Such filtering results in a new set of HRTFs, the internal HRTF, that contain the information necessary for human listening. If, for example, the function 20 $\log_{10}$ is applied to the center frequency of each critical band filter, the frequency domain representation of the internal HRTF becomes a log spectrum that more accurately represents the perception of sound by humans. Additionally, the number of values needed to represent the internal HRTF is reduced from that needed to represent the unprocessed HRTF. An exemplary embodiment of the present invention applies critical band filtering to the set of HRTFs from each individual in the HRTF database **63**, resulting in a new set of internal HRTFs. The process is illustrated in Figure 12, wherein an impulse response waveform **80** shown in Figure 11 is filtered via a critical band filter **81** to produce the internal HRTF **82**.

The application of critical band filtering results in, for example, N logarithmic frequency bands located in the 3000 Hz to 18,000 Hz range. Associated with each of these N frequencies is the level in that band in decibels. In one exemplary embodiment, N=39, the levels are measured with a density of about 15 levels per octave. The entire data base, given S subjects and L locations, is described by 2*S*L*N values and is illustrated in Figure 13. This pre-processing summarizes the more salient perceptual features of the acoustic filtering produced by the head and external ear when a listener hears a sound at a given position in space.

HRTFs obtained from the different subjects and transformed or pre-processed as described above can now be compared and organized so that their similarities and differences can be quantified. One basic method of comparing two or more spectra is the simple Euclidian distance. Euclidian distance is equal to the root-mean-squared (RMS) difference in decibels between the levels measured at the same frequencies in the two or more spectra. For a collection of HRTFs obtained from the right ear of S subjects, we can compare this set by forming a distance matrix having S rows and S columns, in which the entry (i, j) is the distance in decibels between the internally represented HRTF of the "ith" and "jth" individuals. Naturally, the distance measure is symmetric, so the entry (i, j) is equal to the entry (j, i), and the distance between any individual and themselves is zero, so the diagonal entries (i, i), where i=j, are all zero. It is on the basis of the similarities and differences between the processed HRTFs that the database is ordered.

Having explained how the HRTFs are measured and preprocessed, we can now return to the issues raised earlier about how the user of the device selects a particular HRTF from those stored

VSDOCID: <WO_____9725834A2_I_>

in the device. The selection process must ensure that the sound sources appear in their proper spatial position for the individual user. Thus, the first issue to be addressed is whether the entire database of measured HRTFs is sufficiently broad and comprehensive to represent the entire listening population. In one exemplary embodiment, 150 HRTFs were measured from a population in which both genders and a variety of ages and ethnicities were represented.

Statistical tests of this database suggests that 150 HRTFs constitute a set size sufficient for the purposes of the subject invention. These tests were all conducted on a sample consisting of 150 sets measured according to this invention. Three HRTFs from each HRTF set were selected for these comparisons, namely, on the horizon (0 elevation) and at 10, 20, and 30 degrees to the left of straight ahead. It is expected that similar conclusions about stability would apply for other positions. Each of the three HRTFs from each HRTF set consists, for example, of values representing the level of the HRTF, at a plurality, e.g. 39, of different frequencies. The 39 frequencies are spaced equally, on a logarithmic frequency axis, from about 3,000 to about 18,000 Hz. Few listeners (except the very young) can hear sound above 18,000 Hz. The composite spectra obtained over the 3 positions can be regarded as a vector consisting of 117 levels (dB).

To investigate the issue of database size, we constructed different sized sets of HRTFs by drawing them at random from the original group of 150 HRTFs. Set sizes of 20, 40, 60, 80, 100, and 120 HRTFs were constructed. For each of these randomly constructed sets, a single HRTFs is drawn at random and the distance from that individual's HRTF to its nearest neighbor is computed. These random constructions are repeated many times so that the probability of a given distance can be estimated. Figure 22A shows a plot of the cumulative probability of that distance for the various different set sizes. For example, if the set size is 20, then the RMS distance in decibels to the nearest neighbor is less than 2 dB for only about 55% of the individual HRTFs. If the set size is increased to 40 HRTFs, then more than 70% are within 2 dB. As the set size increase to 60, 80, 100, and 120, little incremental advantage is achieved by adding further HRTFs to the database. This analysis demonstrates that the basic differences in HRTFs among different individuals is adequately represented in a database having more than about 100 HRTFs. That is to say, with a raw database containing 100-200 HRTFs there is a very high likelihood that a randomly selected individual would find an HRTF sufficiently close to his/her own so as to properly spatialize sound.

Another way to approach the issue of stability is to compute a significant statistic of the dataset and determine how it changes as we vary set size. From the 150 composite spectra, or vectors, a centroid HRTF is computed. The centroid, itself having 117 levels, is obtained by adding together, for each of the 117 levels, the value representing the level of the HRTF from each of the 150 composite spectra and dividing each sum by the sample size, 150 in the example. If each of the

3NSDOCID: <WO_____9725834A2_I_>

150 composite spectra are treated as a point in a space of 117 dimensions, the centroid is the center of gravity of the set of 150 points.

The Euclidean distance between the centroid and each of the 150 composite spectra (RMS distance in dB) can then be measured. The mean of this distance is about 2.53 dB, and the standard deviation is about 0.76 dB. Figure 22B shows an estimate of a cumulative density function, which is a plot of the probability of an individual being less than a given value, x, from the centroid. As is shown in Figure 22B, the nearest individual in the space was about 1 dB from the centroid; approximately half the sample was within 2.5 dB of the centroid and about 95% were within 4 dB. Also shown in Figure 22B, as a solid line, is a cumulative distribution from a normal or Gaussian distribution with the same mean and standard deviation as the sample, namely, mean = 2.53 dB, standard deviation = 0.76 dB. The data depart somewhat from this theoretical distribution, but the similarity is evident.

Given that these data are reasonably approximated by the normal or Gaussian function, and because the Gaussian distribution is completely described by two parameters, the mean and the standard deviation, the stability of the data is assessed as the number of HRTFs measured is increased or decreased thus defining larger or smaller databases of HRTF subsets, and observing the effect this has on the mean and standard deviation. For this assessment, random subsamples are drawn from the large sample of 150, and the mean and standard deviation of each subsample was calculated. One thousand randomly drawn subsamples for each of five subsample group sizes, namely 5, 10, 20, 40, and 80, were taken. Both a mean and standard deviation of the RMS distance from each of the HRTFs in the subsample to the centroid were computed. The average of the 1,000 means and the average of the 1,000 standard deviations, for each subsample group size were computed. Figure 22C shows the change in the average mean as the number of HRTFs in the subsample increases. Figure 22D shows the change in the average standard deviation as the number of HRTFs in the subsample increases. As can be seen from Figure 22C, the average mean changes by about 10% in value as the subsample group size goes from 5 to 80 HRTFs. The last point on the graph is the mean, 2.53 dB, for all 150 HRTFs. Similarly, referring to Figure 22D, the average standard deviation changes by about 25% in value as the subsample group size goes from 5 to 80 HRTFs. As can be seen from both Figures 22C and 22D there is very little change in the average mean or average standard deviation for subsample group sizes, for example, greater than about 50.

Thus, the two critical statistics of the 150 measured HRTFs are reasonably stable, and we have found that little statistical improvement would be gained by increasing the sample size much beyond 150 samples.

While the preceding has established that the initial database is sufficiently comprehensive to cover an entire population of listeners, it should also be appreciated that not each of the 100-200 HRTFs contributes equally to that result. This is because there is considerable similarity or correlation between certain groups within the entire database. This fact suggests that the raw database can be pruned in some fashion to reduce the total number of HRTFs actually stored in the device. Several different statistical techniques might be used to provide an organization of the database that reveals the underlying correlations. These include one of the variety of multidimensional scaling procedures known in the art. The procedure used in one exemplary embodiment herein was cluster analysis. Specifically, we used a hierarchical agglomerative clustering procedure such as that executed by the statistical program S-Plus™. This procedure uses similarities between the HRTFs as measured in a distance matrix of all 150 HRTFs to produce an ordered tree-like structure to the data. At the highest node of the cluster, all of the HRTFs are contained. Successive nodes contain HRTFs that are similar to each other and different from the remainder, just as biological animals are classified as orders, genera, and species. Figure 15 shows a sample cluster of HRTFs obtained from four subjects. Implicit in this example is the fact that HRTFs of the left and right ear of a single subject are usually nearer in distance than are one person's HRTF to any other person's HRTF. Clustering provides a convenient ordering of the entire database, so that subsets of HRTFs can easily be obtained by selecting similar groups determined by the nodes in the cluster. Those skilled in the art will recognize from this disclosure that other methods of ordering known in the art could be used.

A representative subset of HRTF sets from the entire set of 150 HRTF sets, from which a listener can be matched, is chosen to simplify the matching process. In one embodiment, the HRTF sets within a representative subset are stored for use according to the method of this invention. The greater the number of HRTF sets stored in the device, from which listeners can be matched, the more likely the listener will be matched to an HRTF set similar to the listener's own HRTFs. The disadvantages of having a very large number of HRTF sets stored in the device are that more memory is required to store the HRTF sets, with an accompanying increase in cost of the device. In addition, it would take more time to match the listener with the best-match HRTF set.

In order to balance the competing factors in determining the number of representative HRTF sets to include in the device, we computed the mean minimum RMS distance between an HRTF set randomly selected from the entire measured database of HRTF sets (e.g., 150 HRTF sets) and the representative HRTF set, from the subset of representative HRTF sets chosen to be in the device, nearest to the randomly selected HRTF set, as a function of the number of representative HRTF sets chosen to be included in the device. Figure 22E shows the results from two different algorithms for

NSDOCID: <WO_____9725834A2_I_>

selecting representative HRTF sets. These results are typical of those obtained using a variety of algorithms known in the art which can be used to select representative HRTF sets from the database of HRTFs ordered, for example, by clustering analysis. The illustrated results from both algorithms show the same trends, whether one selects representative HRTFs from the ordered database based on the "popularity" of the representative HRTF (i.e. an HRTF that is closest to the other HRTFs within a given subcluster), or based on the isolation of the representative HRTF (i.e. an HRTF most distant from other HRTFs within a given subcluster). Namely, as the number of representative stored HRTF sets decreases from 150 to 12–15, the mean minimum RMS distance increases slowly. Below about 12–15 stored representative HRTF sets, the mean RMS distance increases much more rapidly. The lowest RMS distance is 1 dB because 1 dB is the average RMS deviation between two measurements of the same individual's HRTF set. Thus, in the present analysis, when an HRTF set randomly chosen from the 150 total HRTF sets is one of the stored HRTF sets, a value of 1 dB is used to represent the RMS distance, not 0 dB. Accordingly, the lowest possible value for the RMS error is 1 dB.

In one embodiment, 25 HRTF sets is the number of representative HRTF sets to be stored in the device, for listeners to select from. This number, 25, is well below the "knee" of the plot in Figure 22E, and is therefore a clearly adequate representative set size, thus balancing the advantages of having a higher number, for example, a closer ultimate match of the listener's HRTF set, and the disadvantages of having a higher number, for example, higher memory cost and a longer matching time for the listener. In one specific embodiment, the listener first chooses from among 5 representative HRTF sets, each representative set representing a set of 5 similar HRTF sets. Once one of the 5 representative sets is selected, the user selects from among the five similar HRTF sets in the set of HRTF sets corresponding to the selected representative HRTF set.

In another preferred embodiment, 15 HRTF sets is the number of representative HRTF sets to be stored in the device for listeners to select from. This number is approximately at the "knee" of the plot in Figure 22E. Having discovered from the aforedescribed statistical analysis of our large ordered database that 15 representative HRTF sets is sufficient to allow the vast majority of the population to select an HRTF set that will allow proper audio spatialization, the 15 representative HRTFs may be selected as follows: the entire database is ordered such that the distance metric (Euclidian distance, RMS distance, etc.) between every HRTF and every other HRTF in the database is known. Thus, in a first step, every HRTF set that is a distance x, e.g., 2, dB away from a particular HRTF set in the database is identified. This identification is made for each HRTF set in the database, and a listing is made of each HRTF set and all of the HRTF sets within x, e.g., 2, dB of it, from the most popular to the least popular HRTF set. The most popular HRTF set is that set in

the database that has the most HRTF sets within x, e.g., 2, dB of it. In a second step, the process of selecting 15 representative sets proceeds by first selecting the most popular HRTF set as a representative HRTF set, and then eliminating every HRTF set that was within x, e.g., 2, dB of the most popular HRTF set from further selection in the database. The next most popular HRTF set, which was not eliminated upon the selection of the most popular HRTF set, is then selected to be the second representative HRTF set, and every remaining HRTF set in the database within x, e.g., 2, dB of this HRTF set is accordingly eliminated. This process is repeated, moving down the list of popularity of HRTF sets that remain in the database. Once 15 representative HRTF sets have been selected, the process may be terminated. Naturally, it will be recognized that fewer or more representative HRTF sets may be selected and that a stringency, i.e., x, of greater than about 1 dB to about 4 dB may be imposed around each of the most popular HRTFs so as to arrive at about 15-25 representative HRTF sets from the entire database of measured HRTF sets. From our statistical analysis, we have found that 15-25 representative HRTF sets is preferred for the considerations provided above.

Once a number of HRTF representative sets have been selected, the user selects the HRTF set that he/she will use in listening to program material by any of several different methods. One procedure is to present, via headphones, sounds filtered by a variety of HRTFs to convey the impression of phantom sounds rotating about the listener's head. The programmed sounds are in fact all chosen from elevations on the horizon. What is generally true of HRTFs is that the variation in the filtered spectrum decreases as elevation increases. That is, the HRTF is generally flatter as the elevation of the sound increases. It is also true that a listener using an HRTF that is very dissimilar to his/her own will tend to hear the phantom sound much higher in elevation than that programmed. Thus, when a listener hears a sound at a lower elevation, it generally means that the listener better appreciates the structure in those HRTFs. Consequently, if one listens to a set of different HRTFs programmed to produce the circle of phantom sounds on the horizon such as that illustrated in Figure 10, the HRTF set producing the lowest apparent elevation will provide the best means to localize sound in the correct spatial location.

Summarizing the foregoing description, the present invention uses HRTF clustering as illustrated in Figure 6B. As discussed above, the present invention collects and stores HRTFs from numerous individuals in the HRTF database **63**. These HRTFs are pre-processed by the HRTF ordering processor **64** which includes an HRTF pre-processor **71**, an HRTF analyzer **72** and an HRTF clustering processor **73**. The HRTF pre-processor **71** processes HRTFs so that they more closely match the way in which humans perceive sound, as described above and further below. The smoothed HRTFs are statistically analyzed, each one to every other one, to determine similarities and

3NSDOCID: <WO_____9725834A2_I_>

differences between them by HRTF analyzer **72**. Based on the similarities and differences, the HRTFs are subjected to a cluster analysis, as is known in the art, and as described above may be "pruned" to arrive at a representative set of HRTFs, by HRTF clustering processor **73**, resulting in a hierarchical grouping of HRTFs. The HRTFs are then stored in an ordered manner in the ROM **65** for use by a listener. From these ordered HRTFs, the listener selects the set that provide the best match via the HRTF matching processor **59**. From the set of HRTFs that best match the listener, the HRTFs appropriate for the location of each phantom speaker are input to their respective logical HRTF processing circuits **10** to **14** of Figure 4.

Having provided a general description of the subject invention, (see Figure 4 above), a specific embodiment thereof is described in greater detail with reference to Figures 23 through 28 hereof.

Referring to Figure 23A, after measuring HRTF sets from a sufficiently large number of individuals, 150 individuals in this example, and performing clustering analysis to select the most representative group of HRTF sets, 15 HRTF sets in this example, the listener is matched to or selects a best-match HRTF set from the 15 most representative HRTF sets. Initially, the HRTF sets of the most representative group of HRTF sets, including the user selected best-match set of HRTFs are stored in an external EEPROM **704** to be accessed during the matching process.

Once the most representative group of HRTF sets is stored in the external EEPROM **704**, an input left **601** and right **602** audio signal, typically from a CD player, VCR, laser disk player, or like source of audio signal are inputted to a circuit **600** for processing of the signals to achieve accurate spatialization of the sound transmitted to the user of the headphones.

The circuit **600** may be custom burned into read only memory on a silicon or like chip, or an off-the-shelf, commercially available chip, such as a Motorola DSP 56007 chip, may be programmed by downloading the appropriate connectors to an electrically erasable programmable read only memory (EEPROM) **710** which reconfigures the DSP 56007 chip each time the chip "wakes up." Referring to Figure 23B, within the circuit **600**, the signals are first routed to a Dolby Prologic® or like decoder **603**, a well defined Dolby Laboratories standard known in the art. The Dolby Prologic® decoder **603** provides four output channels, left **604**, right **605**, center **606**, and surround **607**, intended for loudspeakers located to the front left **608**, front right **609**, front center **610**, and rear center **611** of the listener, see Figure 23C, respectively. Before processing the several output channels, such as the four Dolby Prologic® channels, by filtering with HRTFs, preferably the center channel signal **606** is preprocessed within an early reflection **612** processing circuit, to simulate early reflections that sound waves would encounter in a non-anechoic environment. The output signal of the early reflection processing circuit, the left early reflection **613** and the right early

reflection 614 signals, are preferably added 615, 616 to the left channel signal 604 and to the right channel signal 605, respectively, yielding early reflection processed left 627 and right channel 628 signals.

Referring to Figure 24, one embodiment of this early reflection preprocessing, which is intended to provide a sense of direction and spatial cue, comprises delay tap lines 618, 619 with variable length filter delays 620, 621 and variable magnitude gains 622, 623 for the left and right early reflections, respectively. The length of the delays 620, 621 and the magnitude of the gains 622, 623 can be adjusted, according to the simulated early reflections to be imposed on the signals, by, for example, ambiance 696, theater 624, hall 625, or club 626 control buttons. Means for achieving early reflection processing are known in the art (see U.S. patent No. 5,371,799, incorporated here by reference for this purpose).

Referring again to Figure 23B, next, within the circuit 600, the multiple channels of the signal 627, 628, 606, 607 are processed 663 to create the sensation of phantom loudspeakers by filtering each channel of the signal with a pair of HRTFs, from the best-match HRTF set, corresponding to the intended location for that channel. As noted above, before the HRTF filtering can occur, the user is matched to a best-match HRTF set. The user is preferably matched to a best-match HRTF set, from among the most representative group of HRTF sets of the total database of HRTF sets measured so that when used to process an audio signal the user perceives the corresponding sounds to be localized in the proper spatial positions.

Referring to Figures 28A and 23A, one example of how this matching is accomplished is shown in detail. The HRTF matching process begins by the user pushing an HRTF match mode control button (Ears control) 629, thus entering the HRTF matching mode. This places the user in match mode 1 630. In match mode 1 630, the user may select from one of five clusters of HRTF sets (sets 1–5) in the test bank. Representative HRTFs from each of the five clusters are copied from the external EEPROM 704, which stores the most representative HRTF sets, into the internal RAM 631, see Figure 23A, of circuit 600, for testing. The testing is accomplished by presenting the user, upon the user pushing a noise control 703 button, with sound signals produced by a white noise process 632, Figure 28B, with a linearly decaying envelop 633. The user is first presented with a sound processed by an HRTF 640 corresponding to a first predetermined virtual location, e.g., the front left speaker 634, see Figure 28C, and then the user is presented with a sound processed by an HRTF 641 corresponding to a second predetermined virtual location, e.g., the rear left speaker 635, for each of the representative HRTF sets of the five clusters copied to the RAM 631. The user sequentially listens to each representative set by using the HRTF matching control button 636 to step through the representative HRTF sets 1–5, and ultimately selects which of the sound signals, each generated

INSDOCID: <WO 9725834A2 I >

using a representative HRTF set from one of the five clusters (1–5), which the user perceives as most clearly arriving first from the horizon to the user's front left and then arriving from the horizon to the user's rear left. In this embodiment, the user selects the clearest sound signal by pressing the OK button **637**. The selected sound signal corresponds to the representative HRTF set **638** from one of the clusters of HRTF sets (1–5) which contains the first approximation of the user's best-match HRTF set.

The next step is for the HRTF sets (sets 2.1–2.5 in Figure 28A) from the cluster corresponding to the selected sound signal to be copied **1,000** from the external EEPROM **704** into the internal RAM **631** for further selection by the user. Once again, the user is presented with sound signals produced by a white noise process **632** with a linearly decaying envelop **633** processed first by the HRTF **640** corresponding to the front left speaker **634** and then processed by the HRTF **641** corresponding to the rear left speaker **635**, for each of the five HRTF sets 2.1–2.5 within the cluster corresponding to the previously selected representative set (set 2 in Figure 28A). The user then selects which of the sound signals, each associated with one of the HRTF sets (sets 2.1–2.5 in Figure 28A) of the selected cluster, (2), which the user perceives as most clearly arriving first from the horizon to the user's front left and then from the horizon to the user's rear left. Again, in this embodiment, the user selects this sound signal by pressing the OK button **637**. Upon pressing the OK button **637**, the user has selected the user's best-match HRTF set, for example set 2.2 in Figure 28A, and the user leaves match mode.

In one embodiment, the majority of program material produced by a Dolby Prologic® decoder is contained in the front speaker location (location **610** of Figure 23C). Thus, the device can enable the matching process by producing a transient click-like stimulus e.g., a white noise process **632**, filtered by an HRTF appropriate for the frontal position. Fifteen such HRTFs are used, each appropriate for the set of HRTFs associated with the 15 representative individuals chosen from the entire population of 150 HRTFs. The user selects that HRTF which produces the clearest perception of a phantom sound source located directly in front of the listener. This can enable the matching process to provide a match based on the needs of the application. It should be appreciated that other tests may be more appropriate in other applications, but this simple test is adequate for the current application. For example, if the application requires spatialization of sounds to the sides, HRTFs corresponding to the sides can be used in the matching process.

In one embodiment of this invention, a seat control button **643** is provided which allows the user to select where the user will "sit" in the virtual room with respect to the virtual speakers. For example, the user can select the front-of-the-room **644** seat position, in which case the sound which is to appear from the left **634** and right **645** front phantom speakers will be generated from an HRTF

set (2.2.4 in Figure 28A) measured from an appropriate azimuth angle, i.e., 40 degrees azimuth left or right respectively. In addition, for the front-of-the-room seat position **644**, the front left **634**, front center **646**, and front right **645** virtual speakers will be louder than the rear virtual speakers. In contrast, if a rear-of-the-room seat position **647** is chosen, the front left **634** and right **645** virtual speakers will be generated by an HRTF set (2.2.1 in Figure 28A) measured from a smaller azimuth angle, i.e., 10 degrees azimuth left or right respectively. Additionally, for the rear-of-the-room seat position **647**, the front left **634**, front center **646**, and front right **645** virtual speakers will be softer, than the rear left (surround left) **635** and rear right (surround right) **648** speakers.

Once the user has selected a seat position by pushing a seat control button **643**, 10 HRTFs **651–660**, corresponding to the selected seat position and the best-match HRTF set, are copied from the external EEPROM **704** to the internal RAM **631** for use as digital filters. The 10 HRTFs correspond to the front left, front center, front right, rear left (surround left), and rear right (surround right) virtual speaker locations, with a left and right HRTF for each position **651, 652, 653, 654, 655, 656, 657, 658, 659, 660**. These 10 HRTF sets (**651** through **660**), from the best-match HRTF set (2.2), provide the user with a best-match to the user's own head and pinnae filtering characteristics and simulate the user's selected seat position. Note that for each of the 4 seat positions **644, 661, 662, 647**, 10 different HRTFs are copied to the RAM **631**.

Referring to Figure 25, once the 10 HRTFs (**651** through **660**) are in the internal RAM **631** and available for filtering of the signal, the four standard Dolby Prologic® outputs after early reflection preprocessing, **627, 628, 606, 607**, are fed to the HRTF processing circuit **663**. In one embodiment of the present invention, a fifth channel (second surround channel) **664** may be generated by optionally inverting **665** the single Dolby Prologic® surround channel **607**. This inversion **665** aids in decorrelating the two surround channels. These two surround channels **607, 664** then become rear left (surround left) **607** and rear right (surround right) **664** channels. Accordingly, the surround right channel **664** is identical to the surround left **607** channel, although possibly invented. Each of the five channels (left front **627**, center front **606**, right front **628**, left rear **607**, and right rear **664**) is then split into a right and left channel for filtering by the corresponding HRTFs (**651–660**) stored in the RAM **631**.

Referring to Figure 23A, to prevent loss of HRTFs and other operating mode parameters selected by the user at power-down and power-up, an EEPROM **710** stores all current parameters of the system including current HRTFs, and its stored data is not disturbed by power-up/power-down events. This EEPROM can save, after selection by user, multiple operating mode parameter presets, which can be pulled up by a user by, for example, pushing a button.

3NSDOCID: <WO_____9725834A2_I_>

The HRTF filtering of the 5 left and 5 right channels is accomplished by convolving (or mixing) each channel with the HRTF, from the best-match HRTF set, corresponding to the given location and to the given ear. The convolution of these 10 signals with the corresponding HRTFs produces signals which produce sound corresponding to virtual or phantom speakers at locations corresponding to the locations from which the HRTFs were measured. Once the 10 convolutions are completed, the 5 left signals are summed **666** to generate a summed left signal **668**, and the 5 right signals are summed **667** to generate a summed right signal **669**. These left **668** and right **669** summed signals can be sent directly to a set of headphones for virtual speaker generation. However, additional processing of the summed left **668** and right **669** signals to enhance the effect experienced by the user may be performed. This further processing eliminates the impression of being in an anechoic chamber with the five speakers generating the sounds. Sound in an anechoic chamber does not have the same "fullness" of sound as if the user were in an echoic chamber.

Referring to Figure 23B, to enhance the "fullness" of the sound experienced by the user, bass boost **670** and reverberation **671** processing is preferably performed on the signals before presentation to the user over headphones. These are well known processes in the art. In particular, both the left **668** and right **669** summed output from the HRTF processing may be directed to a bass boost processing block **670**. Referring to **Figure 27**, this circuit **670** comprises, for example, a 100 Hz lowpass filter **672, 673** for each signal, left **668** and right **669**, to produce signals **681** and **682** followed by an amplification **674, 675** of gain G_B for each signal, left and right. The gain G_B can be adjusted, per the user's preference, up or down to adjust the amount of bass boost to the signals by using the bass control button **680**. The left **676** and right **677** outputs of the respective amplifiers are then added to the respective left **668** or right **669** input signal to produce a left bass boosted output **678** and a right bass boosted output **679** signal. The left bass boosted output **678** and right bass boosted output **679** signals are essentially the original signal **668, 669** with an added component comprising G_B times the respective output **681, 682** of the signal through a 100 Hz lowpass filter **672, 673**, thus boosting the bass component of the signals.

Referring to Figure 23B, the left bass boosted **678** and right bass boosted **679** output signals are then added to the output of a reverberation processing circuit **671**, where the inputs **604, 605, 606, 607** to the reverberation processing block are the original four standard Dolby Prologic® or like outputs before any other processing. The reverberation processing **671**, in conjunction with the early reflection processing **612**, provides the "fill" or architectural enhancement that an anechoic representation lacks. Referring to Figure 26, the reverberation processing circuit **671** comprises two all-pole comb filters **683, 684**, in parallel, the summed output of which **692** feeds into two all-pass filters **685, 686** in parallel. The four standard Dolby Prologic® or like outputs are first summed **687**

together and the sum **688** is then inputted to the first comb filter **683** and to the second comb filter **684**. Each all-pole comb filter **683, 684**, as shown in Figure 26, loops the input signal upon itself over and over again with the volume reduced by some fractional amount for each successive loop. The looping has an associated time delay, t = [k] **690**, and gain, G_C **691**, which can be adjusted to suit the user, and are adjusted by the user choosing among a theater **624**, hall **625**, or club **626** setting, with each setting having a unique pairing of length of time delay, t = [k] **690**, and magnitude of fractional gain, G_C **691**. The summed output **692** of the two comb filters in parallel feeds two all-pass filters **685, 686** in parallel. These all-pass filters provide a smearing effect in time to the signal at its input without disturbing the frequency characteristics of the input. The all-pass filters are non-linear phase distorters and remove some of the phase information as a function of frequency. This allows decorrelation of the left **693** and right **694** reverberation outputs, even though the input **692** to the left and right all-pass filters is the same, without disturbing the frequency profile which is embedded in the signal from the HRTF processing. The level of the left **693** and right **694** reverberation outputs is a function of gain, G_R **695**, which is controlled by the ambiance control button **696**.

Referring to Figure 23B, the left **693** and right **694** reverberation outputs are summed **697, 698** with the left **678** and right **679** bass boost outputs, respectively. These summed left **701** and right **702** signals are the left audio out **701** and right audio out **702** signals respectively. The left audio out **701** and right audio out **702** can be sent directly to a set of headphones to provide the listener with the sensation that the audio is originating from virtual speakers positioned according to the seat control selection made by the user. In one embodiment, the headphones are connected via wire to outputs **701** and **702**. In another embodiment, **701** and **702** are signals sent via wireless connection to a set of headphones (see Examples 2, 3, and 4).

Based on the foregoing disclosure, those skilled in the art will appreciate that the method of selecting the best match set of HRTFs from a sufficiently large database of measured HRTFs may be varied considerably, without departing from the principles of this invention. Accordingly, with reference to Figure 29A, by analogy to Figure 28A, with primed reference numerals in Figures 29A and 29B relating to like elements in Figures 28A and 28B, it will be appreciated that a representative set of 15 HRTFs (sets 1-15) may be stored in the test bank. The 15 representative HRTFs used are predicted to accommodate roughly 95% of the population, with respect to variations in the spectral properties of their impulse responses. Again, by analogy to Figure 28A and the foregoing description, the HRTFs are copied, one at a time, from the external EEPROM into the internal RAM of the DSP chip for testing. The user may test these HRTFs by asserting a test signal, see Figure 29B, which will be comprehended by analogy to Figure 28B. A white noise process with a linearly

INSDOCID: <WO 9725834A2 I >

decaying envelope is played from the Center (C) speaker (see figure 28C). The user chooses the HRTF set that best fits the following criteria: (a) the sound source is localized directly in front of the user, and (b) the sound source is localized at the horizon (i.e. on a horizontal plane defined by the user's pinnae). Once the user has identified a set of HRTFs that satisfies these criteria (i.e. has selected a best match HRTF set), the user exits match mode. The seating position can then be adjusted, as described above with reference to Figure 28A, by selecting the 10 HRTFs used by the HRTF processor to localize the virtual sound sources. In this scenario, the user is spared an intermediate step of HRTF matching used in the system shown in Figure 28A.

From the foregoing disclosure, those skilled in the art will also recognize that in an alternative embodiment, rather than matching a user to a representative set of HRTFs wherein the HRTFs used to process an audio signal, for each spatial position, is measured from the same individual, a user can instead be matched to separate representative sets of HRTFs for each spatial position. The user would perform a matching step for each spatial location, wherein a subset of each representative set, selected for the desired spatial position, would be used to process the audio signals. We shall refer to this set herein as a Multi-Position Head-Related Transfer Function or MPHRTF.

In selecting the MPHRTFs, the listener would experience a sound source at each location. The sound source may change for each location depending on the objective criterion at that location. For example, the sound source may be speech for a location in which speech is the main information to be presented. Another may be filtered white noise for those locations that will present ambient noise.

In selecting these HRTFs for each location, a listener would be allowed to choose across multiple sets of HRTFs, where a set of HRTFs is defined to be those recorded from a single subject. This allows the listener to custom develop a "user's set of HRTFs" that best describe his/her localization and perception characteristics at each location to be presented. Furthermore, an interpolation algorithm could generate intermediate locations for the user's set of HRTFs as a mixture of the selected HRTF sets.

Other variations and modifications of these selection schemes will be obvious to those skilled in the art based on this disclosure.

Example 1

In a specific embodiment, the statistical analysis of HRTFs performed by the HRTF analyzer 72, shown in Figure 6B, is performed through computation of eigenvectors and eigenvalues. Such computations are known, for example, using the MATLAB® software program by The

MathWorks, Inc. An exemplary embodiment compares HRTFs by computing eigenvectors and eigenvalues for the set of 2S HRTFs at L * N levels. Each subject-ear HRTF set may be described by one or more eigenvalues. Only those eigenvalues computed from eigenvectors that contribute to a large portion of the shared variance are used to describe a set of subject-ear HRTFs. Each subject-ear HRTF may be described by, for example, a set of 10 eigenvalues.

In this embodiment, the cluster analysis procedure performed by the HRTF clustering processor 73, shown in Figure 6B, is performed using a hierarchical agglomerative cluster technique, for example the S-Plus® program, provided by MathSoft, Inc., based on the distance between each set of HRTFs in multi-dimension space. Each subject-ear HRTF set is represented in multi-dimensional space in terms of eigenvalues. Thus, if 10 eigenvalues are used, each subject-ear HRTF would be represented at a specific location in 10-dimensional space. Distances between each subject-ear position are used by the cluster analysis in order to organize the subject-ear sets of HRTFs into hierarchical groups. Hierarchical agglomerative clustering in two dimensions is illustrated in Figure 14. Figure 15 depicts the same clustering procedure using a binary tree structure.

This embodiment stores sets of HRTFs in an ordered fashion in the ROM 65 based on the result of the cluster analysis. According to the clustering approach to HRTF matching, the present invention employs an HRTF matching processor 59 in order to allow the user to select the set of HRTFs that best match the user. In an exemplary embodiment, an HRTF binary tree structure is used to match an individual listener to the best set of HRTFs. As illustrated in Figure 15, at the highest level 48, the sets of HRTFs stored in the ROM 65 comprise one large cluster. At the next highest level 49, 50, the sets of HRTFs are grouped based on similarity into two sub-clusters. The listener is presented with sounds filtered using representative sets of HRTFs from each of two sub-clusters 49, 50. For each set of HRTFs, the listener hears sounds filtered using specific HRTFs associated with a constant low elevation and varying azimuths surrounding the head. The listener indicates which set of HRTFs appears to be originating at the lowest elevation. This becomes the current "best match set of HRTFs." The cluster in which this set of HRTFs is located becomes the current "best match cluster."

The "best match cluster" in turn includes two sub-clusters, 51, 52. The listener is again presented with a representative pair of sets of HRTFs from each sub-cluster. Once again, the set of HRTFs that is perceived to be of the lowest elevation is selected as the current "best match set of HRTFs" and the cluster in which it is found becomes the current "best match cluster." The process continues in this fashion with each successive cluster containing fewer and fewer sets of HRTFs. Eventually the process results in one of two conditions: (1) two groups containing sets of HRTFs

NSDOCID: <WO_____9725834A2_I_>

so similar that there are no statistical significant differences within each group; or (2) two groups containing only one set of HRTFs. The representative set of HRTFs selected at this level becomes the listener's final "best match set of HRTFs." From this set of HRTFs, specific HRTFs are selected as a function of the desired phantom loudspeaker location associated with each of the multiple channels. These HRTFs are routed to multiple HRTF processors for convolution with each channel.

Example 2

Referring to Figure 7, left **701** and right **702** audio out signals of Figure 23A (or **30** and **31** of Figure 4), can be inputs, for example **754**, of a typical digital signal transmission system known in the art, the output of which, for example **762**, can be inputted to a set of headphones.

Left **701** and right **702** audio out signals (or **30** and **31** of Figure 4) can be outputted in digital or analog format. If outputted in analog format, each signal can be converted to digital format **755**. In a preferred embodiment of this invention, after conversion to digital format, the left and right audio signals are interlaced in time to create a single digital signal **755** which carries both the left and right channel information. For example, the single interlaced digital signal **755** can have a first digital word, e.g., 16 bits, that is a right audio channel word, a second digital word that is a left audio channel word and thereafter alternating between right and left (see Figure 9G). This single digital signal **755** carrying both the left and right audio channel information can then be inputted, for example **755** of Figure 7, to a typical digital signal transmission system.

A standard digital signal transmission system, as shown in Figure 7, typically comprises a transmitting station **751**, a connecting medium called a channel **752**, and a receiving station **753**. The transmitting station **751** can receive an analog signal **754** and convert it to a digital signal **755** or can receive a digital signal **755** directly. Conversion of an analog to a digital signal, for example using an analog-to-digital (D/A) converter **756**, requires the analog signal to be sampled and quantized to the nearest of a number of discrete signal levels. The discrete signal level of the quantized signal is sent to a source encoder **757** where each discrete signal level is converted into a digital representation thereof, typically binary. This representation can consist of digital words, for example 16-bit digital words, wherein each digital word represents the value of a discrete signal level. These digital words can be transmitted sequentially as a serial binary digital bit stream. The binary digital representation is in a particular waveform format, e.g., unipolar or Manchester, and is sent to a modulator **758**, which modulates the signal for transmission over the channel **752**. For instance, the modulator **758** can be a RF modulator, for which the corresponding channel would be air. Alternatively, the channel may be a wire or like transmission means. The receiving station **753** is essentially the inverse of the transmitting station and comprises a demodulator **759**, a source

decoder **760**, and an optional digital-to-analog converter **761**. The output from the receiving station can accordingly be either an analog output **762** or a digital output **763**.

Example 3

Important parameters and design considerations for a digital signal transmission system are bandwidth of the channel, costs of the transmitting and receiving stations, power consumption of the transmitting and receiving stations, and the particular binary waveform chosen for source encoding. Bandwidth is important because it limits the amount of information that can be sent per unit time. The selection of the binary waveform is important because the selection can affect bandwidth and the costs, complexity, and power consumption of the transmitting and receiving stations. This example provides a method for signal transmission that avoids certain problems, discussed below, inherent in known transmission systems for digital signals which enhances the fidelity of the HRTF processed signal of this invention as it is sent to a listener.

Where a receiver, for example, within the receiving station of Example 2, has no clock which is, *a priori*, synchronized to an incoming digital bit stream, the digital bit stream is called an asynchronous signal. When an asynchronous binary format digital bit stream is received, the receiver must, therefore, lock-on to the bit rate in order to generate a clock signal, tied to the bit rate, to enable the receiver to decode the signal. Locking-on to the bit rate can be accomplished by known methods, for example, using a phase-locked loop (PLL). However, there can be difficulties in locking on to the bit rate when receiving digital audio signals represented in binary format, (e.g., two's complement), which are often dominated by repeated strings of contiguous zeroes and/or ones. For example, these strings of contiguous zeroes and/or ones can be encountered with audio signals during moments of silence, or idle patterns. These strings of contiguous zeroes and ones can lead to drifting of the output frequency of the PLL due to an imbalance in the charging and discharging events within the PLL. When the output frequency of the PLL drifts, the PLL can lose its lock, resulting in decoding errors, and thus degradation in the performance of the entire transmission system. In contrast, a binary format digital signal without repeated strings of contiguous zeroes and/or ones would give the PLL a balance of charging and discharging events, allowing the PLL to track the digital signal's frequency more accurately.

Existing solutions for eliminating the drifting of the PLL's lock-in frequency due to repeated strings of contiguous zeroes and/or ones have required additional bandwidth or complicated, expensive hardware. For example, Manchester, or bi-phase-level encoding, commonly used for digital audio signals, eliminates the drifting of the PLL. A Manchester encoded waveform transmits the symbol 1 as a positive pulse for half of the symbol interval, followed by a negative pulse for the

NSDOCID: <WO_____9725834A2_I_>

remainder of the interval; the symbol 0 is conveyed by the same two-pulse sequence but of opposite polarity. Therefore, using Manchester encoding, even with binary format digital signals having repeated strings of zeroes and/or ones, receiver clock timing can be extracted without drifting of the PLL by providing a charging and discharging event for the PLL in the form of a signal transition for each bit received. Accordingly, the Manchester encoding technique allows the PLL to easily lock-on to these regular signal transitions. Unfortunately, Manchester encoding requires about twice the bandwidth of other encoding techniques such as unipolar and bipolar signaling. Additionally, other techniques which have not required as much bandwidth as Manchester have also been employed. However, these techniques are more complicated and therefore more costly to encode and decode.

This example provides a novel solution to these problems and provides a method of efficient carrier stabilization and bit clock embedding. In a specific embodiment, the subject invention includes a novel encoding, transmission, and decoding technique for binary format digital signals. This is particularly advantageous when applied to signals with frequent idle patterns (e.g. digital audio). Advantages of the subject technique include efficient carrier stabilization and bit clock embedding. In addition, this technology provides a low-cost, low power-consumption transmitter/receiver combination for digital signals, including, but not limited to, digital radio frequency (RF) audio signals processed according to this invention to spatialize sound over headphones..

The subject encoding technique can operate on input binary encoded digital signals, typically encoded in two's compliment. The subject technique involves (a) removing the DC component of the input binary encoded digital signal, if present, and, if not already present, adding a small amount of noise to the input binary encoded digital signal, to ensure that each bit location undergoes transitions between the zero and one states, even during idle patterns; (b) inverting, or toggling, every other bit of the binary encoded signal to provide sufficient transitions between adjacent bits to enable the receiver to lock-on to the bit rate and to prevent drifting of the receiver's PLL when long strings of contiguous zeroes and/or ones are present in the input binary encoded digital signal; and (c) encoding a locking bit on the digital signal, for example one locking bit at the start of each word. This locking bit enables the receiver to lock-on to the word pattern of the digital signal, i.e., the position of the digital words within the digital bit stream. In addition to having little or no DC component, the signal should have enough self-noise to ensure frequent transitions from positive to negative values of the signal. Note, if a signal does not have sufficient self-noise, a noise generator is summed with the signal to ensure frequent transitions between positive and negative values for the signal.

The subject encoding technique operates on an input binary encoded digital signal, typically encoded in two's complement. The first step of the subject technique is to remove the DC component of the input binary encoded digital signal, if present. Since the DC component of the signal is removed, this technique is applied to signals where DC coupling is not critical, as in the audio signals of this invention. Since the human ear cannot detect DC sounds, the DC component is not important with respect to digital audio signals. Therefore, this technique is particularly advantageous with respect to processing digital audio signals.

With reference to Figure 8A, the left **701** and right **702** audio out signals (or **30** and **31** of Figure 4) can be outputted in digital or analog format. If outputted in analog format, each signal can be converted to digital format **901**. In a preferred embodiment of this invention, after conversion to digital format, the left and right audio signals are interlaced in time to create a single digital signal **901** which carries both the left and right channel information. For example, the single interlaced digital signal **901** can have a first digital word, e.g., 16 bits, that is a right audio channel word, a second digital word that is a left audio channel word and thereafter alternating between right and left (see Figure 9G). This single digital signal **901** carrying both the left and right audio channel information can then be inputted as shown in Figure 8A.

It is preferred that the DC be removed **902** from the signal after the signal is in digital form **901**, rather than from the analog signal prior to digitization. When one attempts to remove the DC component of an analog signal before digitization, a small DC component is typically introduced into the digital signal during conversion from analog to digital. This DC component introduced into the digital signal is inherent in known analog-to-digital converters and even though small, is undesirable when implementing the subject invention. For instance, during idle patterns of the signal, this residual DC component can cause bit locations to "stick" (i.e. remain in a zero state or a one state) for long periods. This "sticking" can make it possible for the receiver to mistake a "sticking" bit as a locking bit, which as discussed in greater detail below, is a bit which can be encoded on the digital signal and, typically, is always a zero or always a one.

Removing the DC component **902** can be accomplished by many known techniques, for example, by passing the signal through a high pass digital filter. This high-pass filter can be, for example, an infinite impulse response (IIR) high pass digital filter. It is important, when designing the apparatus which is to remove the DC component from the digital signal, that the apparatus does not detrimentally affect the non-DC components of the digital signal. In a specific embodiment, a first-order Butterworth digital high-pass filter, with a 20 Hz corner frequency, is used. In a preferred embodiment, an adaptive filter is used to remove the DC component.

INSDOCID: <WO_____9725834A2_I_>

In a preferred embodiment, an adaptive filter such as that shown in Figure 8B is used to remove the DC component **902** of the input binary encoded digital signal **901**, generated by interlacing in time the digital format representation of left **701** and right **702** audio out signals of Figure 23A (or left **30** and right **31** earphone signals of Figure 4). For clarity we can define the left channel words within **901** as **901l** and the right channel words as **901r**. The input binary encoded digital signal, in a specific embodiment, can be a 16 bit word signal where left and right channel words are interlocked in time such that the first 16 bit word represents the first right channel word and the second 16 bit word represents the first left channel word. Accordingly, each successive 16 bit word alternates between right channel and left channel. In this case, when removing the DC component **902**, it is required to separately remove the DC from the right channel **901r** and the left channel **901l**, due to the independence of the right channel and left channel signals. Therefore, the right **901r** and left **901l** channels are split apart to be operated on independently for removal of the DC component **902**.

For clarity of discussion, the processing of the left channel **901l** will be explained, noting that the right channel **901r** undergoes the same processing independently. Referring to Figure 8B, the digital word of the input signal **901l** is first summed **771** with a tracking constant C[k] **772**, which can initially be zero. The sum **773**, which is also the output of the adaptive filter, then is compared to zero **774**, for example, by observing the sign bit of the word. If the word is less than zero **775**, the tracking constant C[k] **772** is increased by a step size Q_2 **776**, $C[k+1] = C[k]+Q_2$. Alternatively, if the word is greater than zero **777**, the tracking constant C[k] **772** is decreased by a step size Q_1 **778**, $C[k+1]=C[k] - Q_1$. The tracking control variables, Q_1 and Q_2, are dependent upon the amount of gain desired in the adaptation control circuit. This adaptive filter effectively integrates out an average, or DC component, and continually removes it from the source signal.

When the input signal **901l** or **901r** has sufficient self-noise to ensure transitions between positive and negative values even after the DC component is removed, then it is preferred that Q_1 and Q_2 be equal in size. In addition, referring to Figure 8A, if the input signal **901l** or **901r** does not have sufficient self-noise, a noise generator **924** can be used to add in sufficient noise. In a preferred embodiment, if the input signal **901l** or **901r** does not have sufficient self-noise, the adaptive filter of Figure 8B can be used to both remove the DC component and add in sufficient noise, for example, by having $Q_1 = 2Q_2$. In this embodiment, an input signal **901l** or **901r** having a DC component of zero, with no noise, would first be increased by Q_2 to a value of Q_2, then would be decreased by $Q_1 = 2Q_2$ to a value of $-Q_2$, then be increased by Q_2 to a value of zero, and thus repeat through these values. This ensures that each bit location undergoes transitions between the zero and one states, even during idle patterns.

Referring to Figure 9A, 9B, and 9C, the results of a computer simulation of removing the DC component from a gaussian noise source using an adaptive filter, as shown in Figure 8B, are illustrated. In this simulation, a gaussian noise source with a variance of 2.5 mV and a mean of 0.5V is introduced to the adaptive filter. For this simulation, a value for both Q_1 and Q_2 of 0.488 mV is used. Figure 9A shows the original gaussian noise source waveform, Figure 9B shows the value of the tracking constant, C[k], and Figure 9C shows the output waveform of the adaptive filter. These plots are over 2048 samples or about 52 msec. The output waveform clearly has the DC component removed in the latter half of the plot.

Referring to Figures 9D and 9E, the magnitude frequency response of the input gaussian noise waveform and DC shifted output waveform are shown, where Figure 9D is up to $2x10^4$ Hz while Figure 9E shows an expanded view up to 1000 Hz.

Once the DC component has been removed, the next step is to toggle every other bit **903** of the signal. This toggling can be accomplished by known means, for example, by exclusive ORing the signal with a sequence of alternating ones and zeroes, i.e., ...1010...10... The output of an exclusive OR gate is a one if, and only if, only one of the two inputs is a one. Therefore, when an input is exclusive ORed with a zero, the output is the same as the input. However, when an input is exclusive ORed with a one, the output is an inversion of the input. For example, a one exclusive ORed with a one gives an output of zero and a zero exclusive ORed with a one gives an output of one. Referring to Figure 8A, in a specific 16 bit embodiment, every other bit of the encoded signal is inverted by exclusive ORing **903** each word at the signal with 1010101010101010. It should be noted that one could alternatively exclusive OR the signal with 010101...01 and adjust the receiver accordingly. The purpose of this toggling, or inverting of every other bit, is to provide sufficient transitions between adjacent bits to enable a receiver to lock-on to the bit rate. In combination, the removal of the DC component, and subsequent inverting of every other bit, ensures that there will not be repeated strings of contiguous ones or zeroes, and that each bit location is guaranteed to alternate, or flip flop, between the one and zero states, even during idle patterns of the signal.

To illustrate, in a specific embodiment, 24 bit signed two's complement encoding is used. The most significant bit location is the sign bit in the two's complement binary format, where the sign bit is zero for positive and one for negative signal values. Since the DC component of the digital signal has been removed, the digital signal frequently transitions between positive and negative. Therefore, the sign bit location is equally likely to be a one or a zero. Combining the removal of the DC component with the inversion of every other bit ensures each of the remaining 23 bit locations in this 24 bit illustration are also just as likely to be a one or a zero, and there are no repeated strings of contiguous ones or zeroes remaining in the signal.

INSDOCID: <WO_____9725834A2_I_>

By contrast, even when the DC component is removed, if every other bit were not inverted, the 24 bit signal would frequently have positive value words having a string of zeroes in the most significant bits during idle patterns, such as 000000000000000000100101, with only the least significant bits being in a different state than their neighbor bits. Likewise, there would also be many negative value words, with a string of ones in the most significant bits such as 111111111111111110101110, again with only the least significant bits flip-flopping. If the signal, for example due to noise, were such that the signal remains positive or negative for relatively long periods, then these most significant bits can "stick" at a particular value, zero or one, for an equally long period. These "sticking" bits could be mistaken for a locking bit, wherein a locking bit is a bit which can be encoded on the digital signal and, typically, is always a one or always a zero. A locking bit can be located at a certain bit location within a word to allow a receiver to lock-on to the location of the words within the signal by locking on to the locking bit. However, according to the subject invention, after exclusive ORing the signal with 1010...10, 000000000000000000100101 is converted to 101010101010101010001111 and 111111111111111110101110 is converted to 010101010101010100000100. Therefore, after exclusive ORing the signal with 1010 ... 10, it is ensured that the PLL will receive a balanced number of charging and discharging events as well as numerous transitions at the bit rate, thus allowing the PLL to stay locked-on to the bit rate. Additionally, the noise on the signal, sufficient to ensure transitions between positive and negative values of the signal, ensures that no bit will "stick" in a certain state for too long even during idle bit patterns.

A "code violation" within the signal can be used to allow the receiver to determine where each word begins. In order to provide this code violation, a locking bit can be placed at certain locations within the signal. For example, in an audio signal, right and left channel words can be interlocked in time, where each channel can have, for example, 16 bits as shown in Figure 9G. In this case, the locking bit can be located in a certain position of the right channel word, for example, in the least significant bit location. This locking bit then gives the location of the right channel word, as well as the location of the left channel word. This locking bit can be, for example, always a zero or always a one, which allows a receiver to lock on to the locking bit and, therefore, the word pattern of the digital bit stream. In a specific 16 bit word embodiment, after removing the DC and exclusive ORing with 1010...10, each, for example, right word is ANDed **904** with 1111111111111110. This AND operation leaves the first 15 bits of the 16 bit word unchanged and necessarily encodes a zero in the 16th bit location. This guarantees that each right word has as a locking bit, a zero in the least significant bit location, to allow determination of the location of each word in the digital signal at the receiver. It is important to note that it is not necessary for each word or even every other word

to have a locking bit encoded on it. Indeed, a locking bit could be encoded on every third or fourth word. In fact, the limit as to how far apart locking bits can be spaced is determined by the cost and complexity of the receiver to be used.

Once processed as described above, the signal can be transmitted via a wired connection to headphones or through the air. In a specific example, referring to Figure 8A, for wireless transmission, the signal is inputted to a frequency shift keying (FSK) transmitter **905**, such as a RF9901 FSK transmitter chip from RF Micro Devices, which modulates the signal for transmission from a transmitting loop antenna **906**. A corresponding receiving loop antenna **907** receives the incoming FSK modulated signal and sends the signal to a FSK receiver **908**, such as a RF9902 FSK receiver chip from RF Micro Devices, which demodulates the signal. The demodulated signal can then be inputted to conventional two transducer headphones for listening.

The receiver should be able to lock on to the bit rate and then lock on to the locking bit in order to decode the signal. Referring to Figure 9F, the receiver can comprise a phase lock loop **815**, which provides a master clock **804** and aligns the clocking bits with the data bits provided from, for example, an RF demodulator. The receiver can further comprise a state machine **800**, which can be the center of the timing for the receiver, and can also perform a number of operations including: clocking functions for the D/A converter, reclocking of the data delivered to the D/A, and control lines for master reset. The state machine can provide a serial clock **805**, SCLK, a left/right clock **806**, L/R CLK, and data **803**, SDATA, to a D/A converter. The state machine **800** can, for example, be a free running eight bit counter. Where the signal is transmitted wirelessly, the state machine **800** receives the RF data **801** (RF Digital) and inverts the bits which were inverted prior to transmission, by exclusive ORing RF Digital **801** with a clocking signal Q3 **802** which has a frequency one half of the bit rate (or 1/16 of the master clock). The data stream can then be latched to produce a strong, clean data bit stream, **803** (SDATA), to present to the D/A converter.

The locking bit is encoded on the incoming data stream, RF Digital **801**, to allow the receiver to maintain word lock. The locking bit can be, for example, always 0 (logic level low) in the least significant bit of the digital data word. The state machine **800** looks for the locking bit during a window of time, the locking bit window **808**, to determine if lock is being maintained. If a 0 is present, no action is taken; however, if a 1 is detected, the state machine **800** resets itself via its reset control line **809**. After resetting, the state machine **800** can, for example, start over at a new data position and the process continues until lock is regained. It should be understood that the locking bit could always be 1 and then the state machine would reset upon detecting a 0 during the locking bit window **808**.

NSDOCID: <WO_____9725834A2_I_>

In a specific embodiment, returning to Figure 8A, the demodulated signal output from the FSK receiver **908**, called RFDIG **801**, is in the same binary format as the signal which entered the FSK transmitter **905**. In order to decode the signal, it is inputted to a phase-locked loop (PLL) **815** and also inputted to an exclusive OR gate **917** to be exclusive ORed with 1010...10. The PLL **815** is able to lock on to the frequency of the bit rate due to sufficient bit transitions provided by the exclusive ORing of the signal with 1010 ... 10 prior to transmission, which provides a strong frequency component at the bit rate and provides the PLL **815** a balanced number of charging and discharging events. The output of the PLL **815** is the master clock **804**, MCLK, which has a frequency eight times the bit rate. The MCLK is inputted to a divide-by-eight state machine **912**, with the output thereof, at a frequency equal to the bit rate, fed through a feedback loop **913** to the PLL **815** and fed to latch **916**. Additionally, MCLK **804** is inputted to a state machine **800** which generates clock signals at MCLK/2 (or QO)**810**, MCLK/4 (or Q1)**811**, MCLK/8 (or Q2)**805**, MCLK/16 (or Q3)**802**, MCLK/32 (or Q4)**812**, MCLK/64 (or Q5)**813**, MCLK/128 (or Q6)**814**, and MCLK/256 (or Q7)**806**, wherein MCLK/2 means a clocking signal at the MCLK frequency divided by 2, etc. Figure 9G shows how these clock signals align with each other, the input signal RF digital **801**, the output of exclusive OR gate **917**, XOR output **816**, and the output of latch **916**, SDATA **803**.

Figure 9G shows two 16-bit words, right channel word D15, D14, ... , D0, and left channel word D15, D14, ... , D0, from a digital bit stream, RFDIG **801** in Figure 8A. Note, these two 16-bit words could be considered one 32-bit word. In this embodiment, the first D15, D14, ... , D0 can be a right channel word and the next D15, D14, ... , D0 can be a left channel word. MCLK/8 (or Q2 **805**) is referred to herein as SCLK, the data clock at twice the bit rate, which can be used to determine the state, one or zero, of each bit. To lock on to the locking bit, located at D0 of the right channel word, an eight input NAND gate **915** with inputs NOT Q7 **817**, Q6 **814**, Q5 **813**, Q4 **812**, Q3 **802**, NOT Q2 **818**, NOT Q1 **819**, and a bit value from latch **916**, SDATA **803** after inversion, **922**, is used. Latch **916** can delay each bit for one cycle of MCLK/4, or one-half the duration of a bit. Therefore, the output from latch **916**, SDATA **803**, is delayed with respect to the output of the exclusive OR **917**, by one-half the duration of a bit. This latching and delay allows the bit to be clean and strong during the locking bit window **808**. Figure 9G illustrates the alignment of SDATA **803**, and the various clock signals when the state machine is in lock with the locking bit.

However, before attaining lock on to the locking bit, the bit value during the locking bit window **808**, one or zero, from latch **916** is the bit value of Dn, which is any one of D15, D14, ... , D0, D15, D14, ..., D0 from either the left or right channel word as shown in Figure 9G. The bit value of Dn is obtained by Exclusive ORing **917** RFDIG **801** with Q3 **802**. Exclusive ORing **917**

Q3 802 with RFDIG 801 inverts the previously inverted bits to generate a data signal, XOR output 816, which is a replica of the original binary coded format signal 901 with the DC removed. Q3 802 is synchronized with RFDIG 801, by locking on to the bit rate. After the PLL 815 has locked on to the bit rate, the locking bit is located by first resetting the state machine at a random position within the two 16 bit word cycle. If the output 921 of the NAND gate 915, after inversion by inverter 920, is a zero, then the selected bit is a one and therefore not the locking bit. Alternatively, the inverted NAND gate 915 output 921 will be one only when the inverted bit 922 from SDATA 803, is a one, corresponding to the bit from SDATA 803, the locking bit, being a zero. The inverted NAND gate 915 output, 921, can only be a one if the inverted bit 922 from SDATA 803 is a one at the same time that NOT Q7 817 is a one, Q6 814 is a one, Q5 813 is a one, Q4 812 is a one, Q3 802 is a one, NOT Q2 818 is a one, and NOT Q1 819 is a one, based on the inputs to the NAND gate 915. As can be seen from Figure 9F, this only occurs at the D0 bit location of the right channel word. Therefore, if Dn (n≠0) is arriving when D0 should arrive, then the inverted NAND 915 output 921 remains zero until Dn eventually becomes a zero.

If, in Figures 8A and 9G, Dn is a one, then the inverted NAND gate 915 output 921 is zero, and the state machine 800 can be instructed to reset to the bit following Dn, namely Dn+1. Since each bit location from D15, D14, ..., D0, D15, D14, ..., D0 is guaranteed to alternate between one and zero, except the locking bit, D0 of the right channel word which is always zero, the state machine can quickly lock on to the location of the locking bit. In this synchronized state, lock-on to the locking bit has been achieved. The need to locate the locking bit is why it is imperative that each of the other bit locations are guaranteed to switch to a one state some time in the bit stream such that no other bit location remains in the zero state long enough to be mistaken as the locking bit.

Example 4

In an embodiment such as described in Example 2 or Example 3, if the digital signal is wirelessly transmitted through the air, for example from an FSK transmitter to a FSK receiver, the receiver can be located in a remote unit while the transmitter can be located in a base unit. The base unit can, for example, comprise the HRTF processing circuitry including DSP chip 600, EEPROM 710, and External EPROM 704, such as exemplified in Figure 23A, as well as the signal processing circuitry 901, 924, 902, 903, 904, FSK transmitter 905, and transmitting loop 906, such as exemplified in Figure 8A. The remote unit can, for example, comprise receiving loop 907, FSK receiver 908, PLL 815, state machine 800, NAND gate 800, and associated circuitry exemplified in Figure 8A, as well as input means for HRTF matching control 636, OK control 637, Noise control 703, Bass control 680, Ears control 629, Seat control 643, Ambience control 696, Theater control

NSDOCID: <WO_____9725834A2_I_>

624, Hall control 625, and Club control 626. Alternatively, the input means for the aforementioned control functions can instead be located in the base unit. The headphones can be plugged into the base unit or the remote unit to allow the headphone user to listen to the audio signal. The wireless transmission of the signal from the base unit to the remote unit allows the listener a greater range of motion than if connected to the base unit by wire. If the input means for the control features are in the remote unit, it is preferred to have some means for the remote unit to send information to the base unit.

In a specific embodiment, the remote unit sends information to the base unit, for example, by an infra-red (IR) signal. Specifically, the remote unit has input means, for example, buttons, for the listener to enter, for example, club 626, hall 625, theater 624, ambience 696, seat control 643, ears control 629, bass control 680, noise control 703, OK control 637, and/or HRTF matching 636 signals. These command signals are transmitted to the base unit by, for example, IR.

In order for the remote unit to determine if the base received the IR signal, the base sends a return signal from the base unit to the remote unit, in response to receiving the IR signal from the remote unit. In a preferred embodiment, the subject invention encodes a tag bit on the RF digital audio signal which, when received by the remote unit, indicates receipt, by the base unit, of an IR signal from the remote unit.

This tag bit is a bit encoded similarly to the locking bit. For example, if the locking bit is encoded in the least significant bit location of the right channel word of the audio signal, then the tag bit is, for example, encoded in the least significant bit location of the left channel word of the audio signal. In a preferred embodiment, the tag bit is encoded, as a default value, opposite to the value of the locking bit. For instance, if the locking bit is encoded as one, or a zero, then the tag bit will be encoded, as a default value, as a zero, or a one, respectively. In a specific embodiment where the locking bit is encoded as a zero, the default value of the tag bit can thus be a one and can therefore be encoded by ORing each left channel word with 0000000000000001.

In operation, the receiver in the remote unit interprets a one in the tag bit location to mean that no IR signal has been received by the base unit. When the base does receive an IR signal from the remote unit, the base unit encodes a zero value in at least one consecutive tag bit location by ANDing at least one left word with 11...10 instead of Oring with 00...01. In a preferred embodiment, a zero value is encoded for eight consecutive tag bits to reduce the effects of noise, i.e. bit errors.

The state machine 800 monitors the tag bit location, which is known relative to the locking bit location. In a preferred embodiment, the locking bit is encoded in the least-significant bit location of the right channel word and the tag bit is encoded in the least significant bit location of the left

channel word. In this embodiment, the receiver of the remote unit monitors the tag bit much like it monitors the locking bit. For example, an additional eight input NAND gate similar to NAND gate **915** having inputs Q7 **806**, Q6 **814**, Q5 **813**, Q4 **812**, Q3 **802**, NOT Q2 **818**, NOT Q1 **819**, and a bit value from latch **916**, SDATA **803**, after inversion, **922**, is used. Note, these are the same inputs for monitoring the locking bit location, except NOT Q7 **817** is replaced with Q7 **806**. Figure 9F illustrates the alignment of SDATA **803**, and the various clock signals when the state machine is in lock with the locking bit.

If the inverted output of the NAND gate is a zero, then the tag bit is a one and therefore no IR signal has been received by the base. Alternatively, the inverted output of the NAND gate will be a one only when the inverted bit **922** from SDATA **803** is a one, corresponding to the bit from SDATA **803**, the tag bit, being a zero. A zero value for the tag bit signifies the base unit has received an IR signal from the remote.

The state machine **800** only looks for the tag bit during a small window in time, the tag bit window **820**, after a command is sent via the IR link. The remote clears the tag bit latch, transmits the command word over the IR, and then watches for a zero bit to be latched onto the tag bit control line. If a zero is latched, then the command was received by the DSP, the base; if a one is latched, then the command was not received and no action is taken by the remote unit. When a one is latched and no action is taken by the remote, the user would be required to press the command button again and resend the command over the IR link. Once the receiver locks on to the locking bit, the location of the tag bit will then be known.

It should be understood that the examples and embodiments described herein are for illustrative purposes only and that various modification or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and the scope of the appended claims.

NSDOCID: <WO_____9725834A2_I_>

Claims

1. A method for processing a signal comprising at least one channel, wherein said at least one channel has an audio component, wherein said method allows a user of headphones to receive at least one processed audio component and perceive that the sound associated with each audio component has arrived from one of a plurality of positions, determined by said processing, wherein said method comprises the steps of:

(a) receiving the audio component of each said at least one channel;

(b) selecting, as a function of a user of headphones, a best-match set of head related transfer functions (HRTFs) from a database of sets of HRTFs;

(c) processing the audio component of each said at least one channel via a corresponding pair of digital filters, said pairs of digital filters filtering said audio components as a function of the best-match set of HRTFs, each corresponding pair of digital filters generating a processed left audio component and a processed right audio component;

(d) combining said processed left audio component from each said at least one channel of the signal to form a composite processed left audio component;

(e) combining said processed right audio component from each said at least one channel of the signal to form a composite processed right audio component;

(f) applying said composite processed left and right audio components to headphones, to create a virtual listening environment wherein said user of headphones perceives that the sound associated with each audio component has arrived from one of a plurality of positions, determined by said processing.

2. The method, according to claim 1, wherein said database of sets of HRTFs is generated by measuring and recording sets of HRTFs from a representative sample of the listening population.

3. The method, according to claim 1, wherein each position of said plurality of positions is predetermined and corresponds to one of said at least one channel.

4. The method according to claim 3, wherein, after the step of selecting a best-match set of HRTFs, said method further comprises the step of selecting a position subset of HRTFs from the best-match set of HRTFs, each of the selected HRTFs of said subset of HRTFs being selected so as to correspond to a virtual position closest to one of said predetermined positions so that the

user of said headphones perceives that the sound associated with each said at least one channel originates from or near to said corresponding predetermined position.

5. The method according to claim 1, further comprising any one or all of the following steps:

(a) processing the audio component of at least one of said at least one channel of the signal via a bass boost circuit prior to processing said audio component of said at least one channel via the pair of digital filters;

(b) prior to applying the composite processed left and right audio components to the headphones, further processing the composite processed left audio component and the composite processed right audio component via an ear canal resonator circuit.

6. The method according to claim 1, wherein said audio component of each said at least one channel of the signal is processed such that said predetermined positions are specified by a Dolby Pro Logic® audio component.

7. The method, according to claim 1, further comprising the steps of:

(a) collecting a database of measured HRTFs;

(b) ordering said database so that a representative subset of the entire collection of HRTFs is obtained and stored in storage means; and

(c) selecting a best-match set of HRTFs from said storage means such that a user performing said selecting perceives audio signals processed using said best-match set of HRTFs in the proper spatial positions.

8. The method of claim 7 wherein said database is ordered by clustering said measured HRTFs.

9. The method of claim 7 wherein said representative subset comprises between 15 and 25 HRTF sets.

INSDOCID: <WO_____9725834A2_I_>

10. The method of claim 8 wherein said database comprises S*L*2 spectra, with
L = the number of locations measured; and
S = the number of difference subjects measured, wherein
$16 < S < 200$.

11. The method according to claim 8, wherein the step of matching the user to the best-match HRTF set via HRTF clustering further comprises the steps of:

(a) performing cluster analysis on the database of HRTF sets based on the similarities among the HRTF sets to order the HRTF sets into a clustered structure, wherein there is defined a highest level cluster containing all the sets of HRTFs stored in the database, wherein each cluster of HRTF sets contains either one HRTF set, only HRTF sets which have no statistical difference between them, or a plurality of sub-clusters of HRTF sets;

(b) selecting a representative HRTF set from each one of a plurality of sub-clusters of the highest level cluster of HRTF sets;

(c) selecting a virtual target subset of HRTFs from each representative HRTF set, wherein each position subset of HRTFs is associated with a predetermined virtual target position;

(d) providing, to the user, a plurality of sound signals, each of said plurality of sound signals being filtered by one of said plurality of position subsets of HRTFs;

(e) selecting, by the user, one of said plurality of sound signals as a function of appropriate sound spatialization to said predetermined virtual target position, the selected sound signal corresponding to the best-match cluster, wherein the representative HRTF set of the best-match cluster defines the best-match HRTF set.

12. The method according to claim 11, wherein each selected representative HRTF set is a centroid or popular HRTF which most exemplifies the similarities between the HRTF sets within the sub-cluster of HRTF sets from which the representative HRTF set is selected.

13. The method according to claim 11, wherein each selected representative HRTF is an isolated HRTF which is most different from the HRTF sets within the sub-cluster of HRTF sets from which the representative HRTF set is selected.

14. The method according to claim 11, wherein the step of matching the listener to the best-match HRTF set via HRTF clustering further comprises the steps of:

(a) after selecting, by the user, one of said plurality of sound signals as a function of said predetermined virtual target position, selecting a representative HRTF set from each sub-cluster of the best-match cluster;

(b) selecting a subset of HRTFs from each representative HRTF set of each sub-cluster of the best-match cluster, wherein each subset of HRTFs is associated with a predetermined virtual target position;

(c) providing, to the user, a plurality of sound signals, each of said plurality of sound signals filtered with one of said plurality of subsets of HRTFs corresponding to the plurality of sub-clusters of the best-match cluster;

(d) selecting one of said plurality of sound signals as a function of a predetermined virtual target position, the selected sound signal corresponding to the best-match cluster, wherein the representative HRTF set of the best-match cluster defines the best-match HRTF set;

(e) repeating steps a through d until the best-match cluster contains only one HRTF set or contains only HRTF sets which have no statistical difference between them.

15. A device for processing a signal comprising at least one channel, wherein each said at least one channel has an audio component, wherein said device processes each audio component such that a user of headphones can receive the processed audio component from each said at least one channel and perceive that the sound associated with each audio component has arrived from one of a plurality of positions, said device comprising:

(a) at least one pair of digital filters, each pair of digital filters receiving an audio component and applying a pair of head related transfer functions (HRTFs) to said audio component, the HRTFs being determined as a function of a user of the headphones from a database of sets of HRTFs, each pair of digital filters generating a left signal and right signal;

(b) a first combining circuit combining the left signals for each said at least one channel to form a left output signal; and

(c) a second combining circuit combining the right signals for each said at least one channel to form a right output signal, the left and right output signals, when applied to the headphones, creating a virtual listening environment wherein a user of said

3NSDOCID: <WO_____9725834A2_I_>

headphones perceives that the sound associated with each audio component has arrived from one of a plurality of positions, determined by said processing.

16. The device according to claim 15, further comprising any one or more of:

(a) a bass boost circuit coupled to at least one pair of digital filters, the bass boost circuit increasing a low frequency energy of a signal input to the bass boost circuit;

(b) an ear canal resonator circuit coupled to the left and right output signals; and

(c) a reverberation circuit coupled to at least one of said at least one channel, a first output and a second output of the reverberation circuit being coupled to a respective one of the first and second combining circuits.

17. A method for producing sound over headphones that is accurately spatialized for a given user of the headphones which comprises:

(a) providing said user with a control device which controls a PROM programmed with a database of representative HRTFs sets amenable to selection by said user of a best-match HRTF set;

(b) transferring and storing said best-match HRTF set to RAM linked to a DSP; and

(c) processing an audio signal by said DSP using said best-match HRTF set and transmitting said processed audio signal to said user for perception.

18. The method of claim 17 wherein said processing comprises decoding said signal into a plurality of signals prior to using said best-match HRTF set and, in addition to said processing using said best-match HRTF set, optionally processing components of said plurality of signals by a method selected from the group consisting of early reflection processing, reverberation processing, bass boost processing, and any combination thereof.

19. The method according to claim 18 wherein said selection of said best-match HRTF set comprises transmitting sound via headphones to a user from a main processing device programmed with a plurality of HRTF sets which are representative of major clusters of HRTF sets in a database of HRTF sets measured from a sufficient number of individuals in the general population such that a statistical analysis of the measured data reveals that there would be little incremental enhancement in the fidelity of sound spatialization if a greater number of representative

ISDOCID: <WO_____9725834A2_I_>

HRTF sets were used to program said processing device, and allowing the user to identify a first approximation of a best-match HRTF set by localizing sounds in pre-determined virtual locations.

20. The method according to claim 19 wherein said database of representative HRTFs is selected from a database of measured HRTF sets, generated by measuring the individual HRTF sets of at least sixteen individuals wherein said measuring is achieved using a single robot-arm positioned sound source.

21. A device for producing sound over headphones that is accurately spatialized for a given user of the headphones which comprises:

(a) a peripheral control device which controls a PROM programmed with a database of representative HRTFs sets from amongst which said user is able to select a best-match HRTF set; and

(b) a Random Access Memory (RAM) resident within a main processing device which is programmed with said best-match HRTF set.

22. The device according to claim 21 comprising a means for wired or wireless transmission of sound processed by said main processing device programmed with said best-match HRTF set.

23. The device according to claim 22 wherein said sound is a digital signal and said means for wireless transmission is a digital processing means comprising:

(a) a filtering means for removing the DC component from said digital signal;

(b) a first inverting means for inverting every other bit of said digital signal; and

(c) an encoding means for encoding a locking bit into said digital signal.

24. The device according to claim 23 wherein any one or more of the following apply:

(a) said digital signal is a binary digital signal;

(b) said filtering means is an adaptive filter;

(c) said filtering means is a high-pass filter;

(d) said first inverting means is an exclusive OR gate having as inputs said digital signal and a digital bit stream comprising alternating ones and zeroes (...101010...); and

3NSDOCID: <WO_____9725834A2_I_>

(e) said encoding means is an AND gate having as input said digital signal and a repeating sequence of (...111111...10...), wherein said AND gate encodes a zero as a locking bit every n^{th} bit, where n is an integer.

25. The device according to claim 24 wherein said digital signal is comprised of digital words, wherein said locking bit is encoded into the least significant bit location of each digital word into which it is encoded.

26. The device, according to claim 25 wherein said locking bit is encoded into each digital word as the terminal bit of each said digital word into which it is encoded.

27. The device, according to claim 24 further comprising:
(a) a transmitting means for transmitting said digital signal; and
(b) a receiving means for receiving said digital signal.

28. The device, according to claim 27, wherein said receiving means comprises:
(a) a first locking means for locking onto the bit rate of said received digital signal;
(b) a second locking means for locking onto the locking bit of said received digital signal; and
(c) a second inverting means for inverting said previously inverted bits.

29. The device, according to claim 28, wherein any or all of the following apply:
(a) said first locking means is a phase locked loop;
(b) said second locking means is a state machine; and
(c) said transmitting and receiving means are wireless.

30. A device for rapidly and accurately generating a database of HRTF sets based on measurements from a large number of individuals comprising:
(a) a single, robot-arm positioned sound source;
(b) a robot-arm for positioning said single sound source;
(c) a measurement control system; and
(d) transducers for measuring sound and distortions thereof as it is received at each ear of an individual whose HRTF sets are being measured, after being generated by

said single sound source at various locations about the individual wearing said transducers.

31. The device of claim 30 wherein said transducers are positioned at the entrance of the outer ear canal of the individual whose HRTF sets are being measured.

32. A device for spatializing sound over headphones which comprises:

(a) a means for storing a representative set of HRTFs selected from a database of measured HRTFs;

(b) a means for a user to select a set of HRTFs from said means for storing said representative set of HRTFs; and

(c) a means for processing audio signals using said set of HRTFs selected by the user such that the user perceives the corresponding sounds to be localized on the proper spatial positions;

wherein said database of measured HRTFs comprises S*L*2 spectra, with

L = the number of locations measured, and

S = the number of difference subjects measured, wherein

$16 < S < 200$.

33. The method according to claim 17 wherein said signal is a digital signal and said transmitting comprises:

(a) removing the DC component of said digital signal if present;

(b) inverting every other bit of said digital signal; and

(c) encoding a locking bit into said digital signal.

34. The method according to claim 33, wherein any one or more at the following apply:

(a) said digital signal is a binary digital signal;

(b) said removing of said DC component is achieved by adaptive filtering;

(c) said removing of said DC component is achieved by high-pass filtering;

(d) said inverting of every other bit of said digital signal is accomplished by exclusive ORing said digital signal with a digital bit steam comprising alternating ones and zeroes (...101010...);

INSDOCID: <WO_____9725834A2_I_>

(e) said encoding of a locking bit into said digital signal is achieved by encoding said locking bit in a certain bit location of every n^{th} word comprising said digital signal, wherein n is an integer; and

(f) said encoding of a locking bit into said digital signal is achieved by encoding said locking bit at every n^{th} bit of said signal wherein said locking bit is always a one or always a zero and wherein n is an integer.

35. The method, according to claim 17, further comprising the steps of:

(a) transmitting said digital signal; and

(b) receiving said digital signal to produce a received digital signal.

36. The method according to claim 35, wherein said receiving step comprises:

(a) locking onto the bit rate of said received digital signal;

(b) locking onto the locking bit of said received digital signal; and

(c) inverting the previously inverted bits.

37. The method, according to claim 36, wherein any one or more of the following apply:

(a) said locking onto said bit rate of said received digital signal is accomplished by a phase locked loop; and

(b) said locking onto said locking bit of said received digital signal is accomplished with a state machine.

38. A storage means encoded with a database of HRTFs such that HRTFs appropriate for a particular individual may be retrieved from such storage means to act as a filter in digital processing of an audio signal transmitted to headphones for accurate sound spatialization.

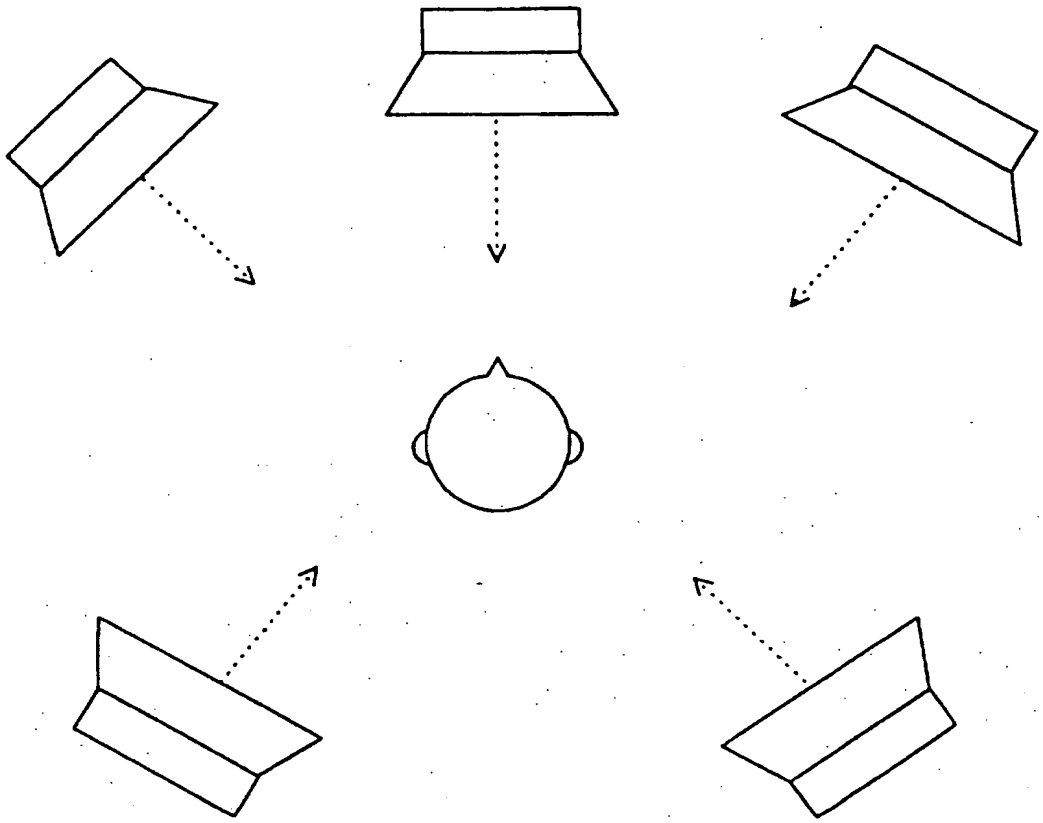


FIG. 1

NSDOCID: <WO_____9725834A2_I_>

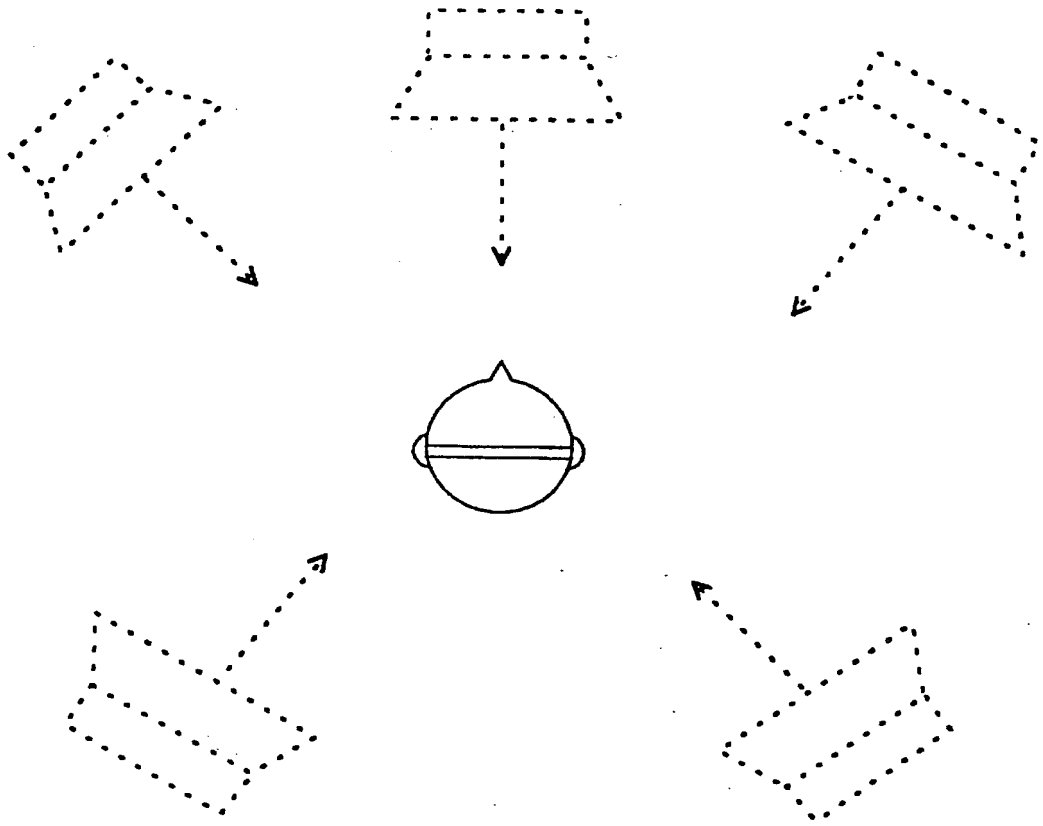


FIG. 2

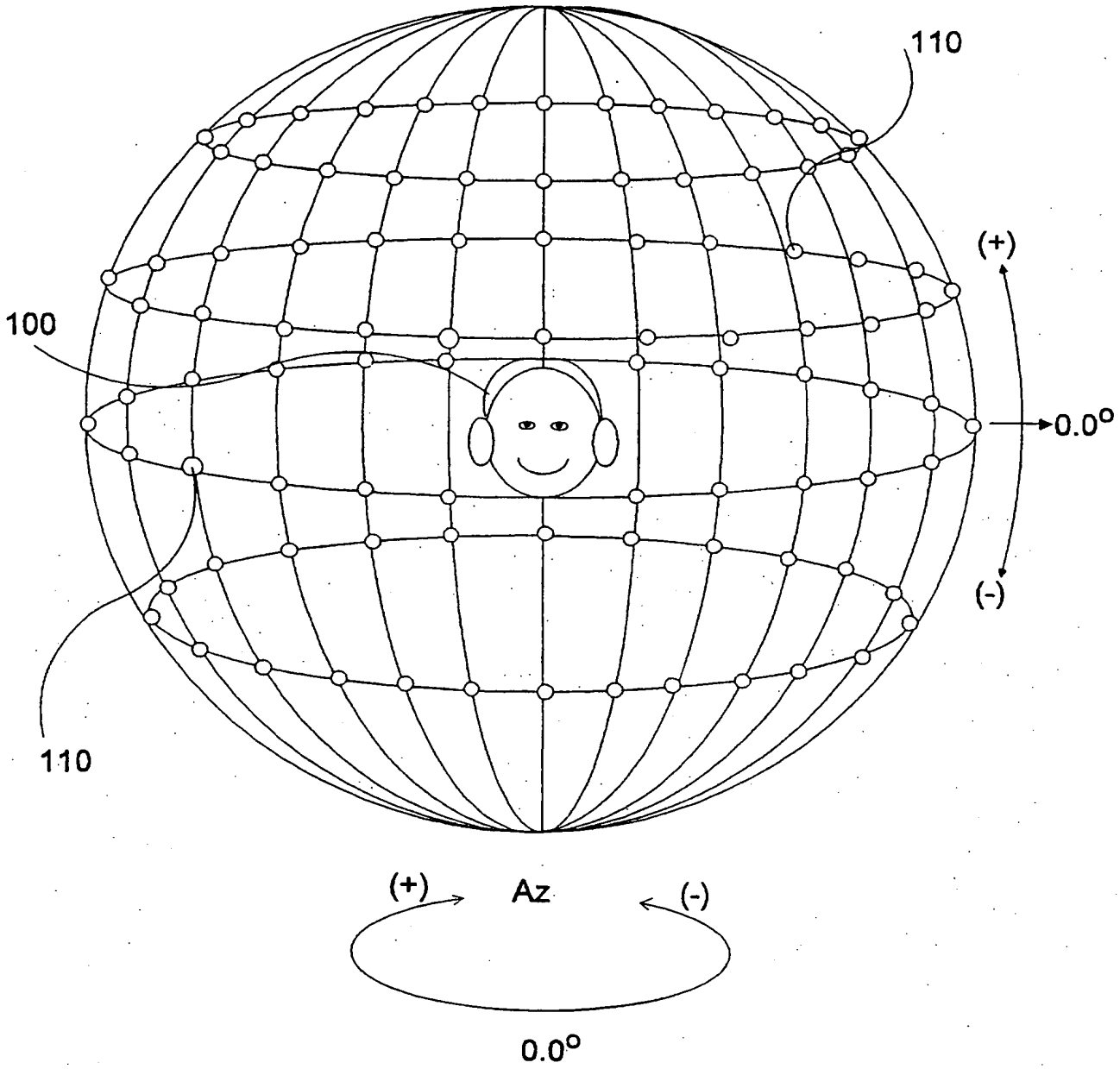


FIG. 3A

INSDOCID: <WO_____9725834A2_I_>

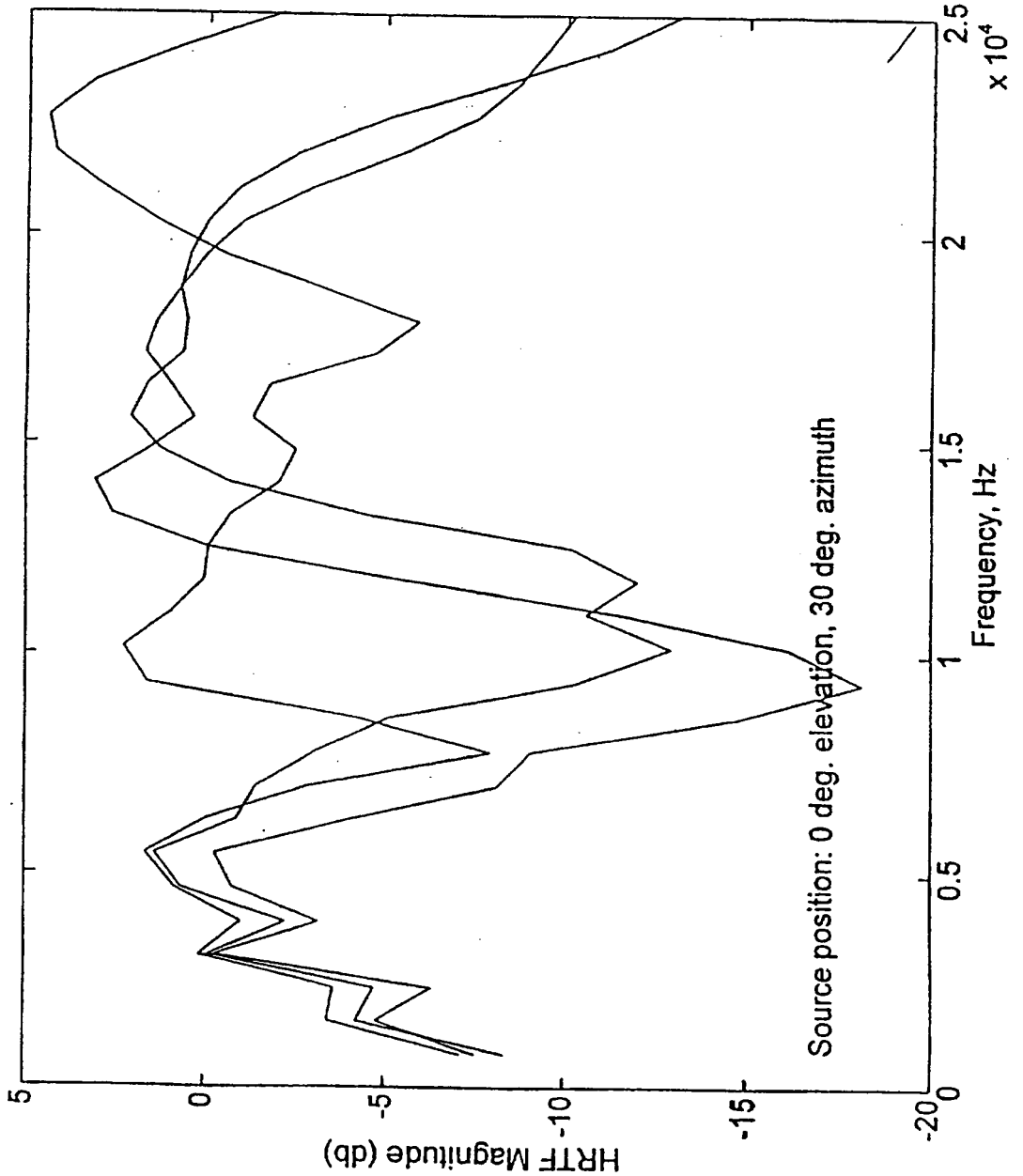

HRTF Magnitude (db)
5
0
-5
-10
-15
-20
0
0.5
1
1.5
2
2.5
x 10^4
Frequency, Hz
Source position: 0 deg. elevation, 30 deg. azimuth

FIG. 3B

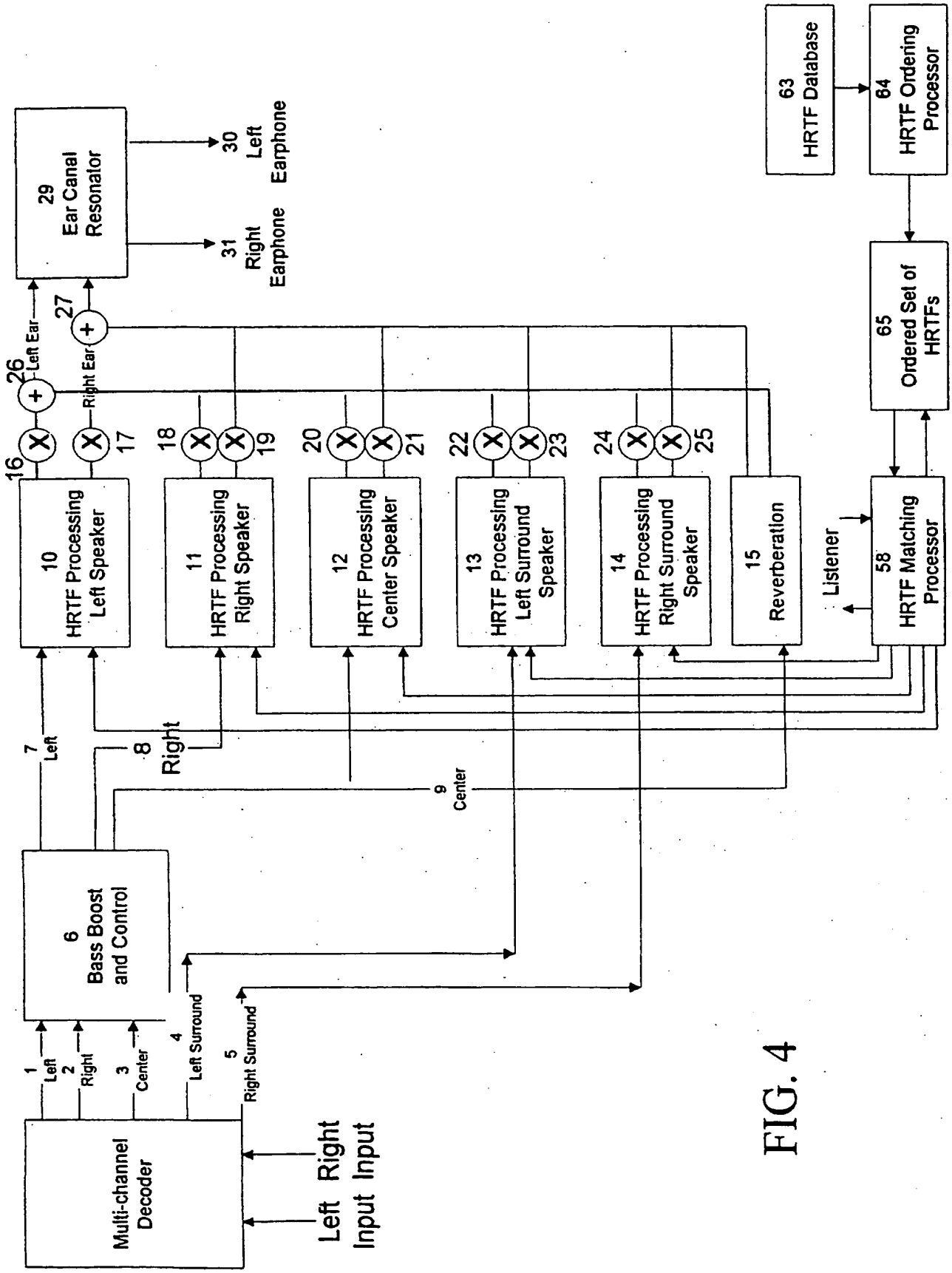


FIG. 4

INSDOCID: <WO_____9725834A2_I_>

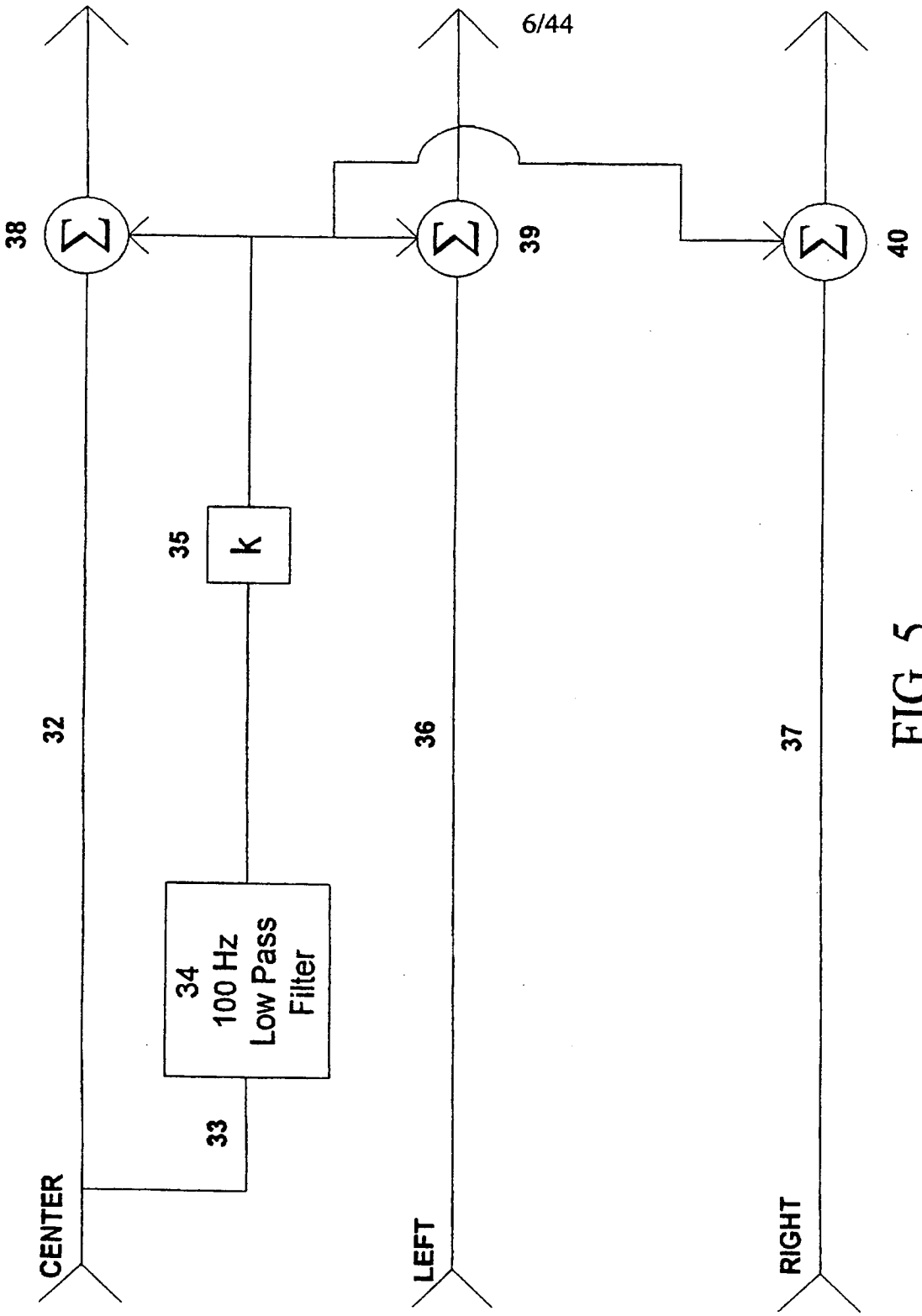

38
39
40
35
k
32
36
37
34
100 Hz
Low Pass
Filter
33
CENTER
LEFT
RIGHT

FIG. 5

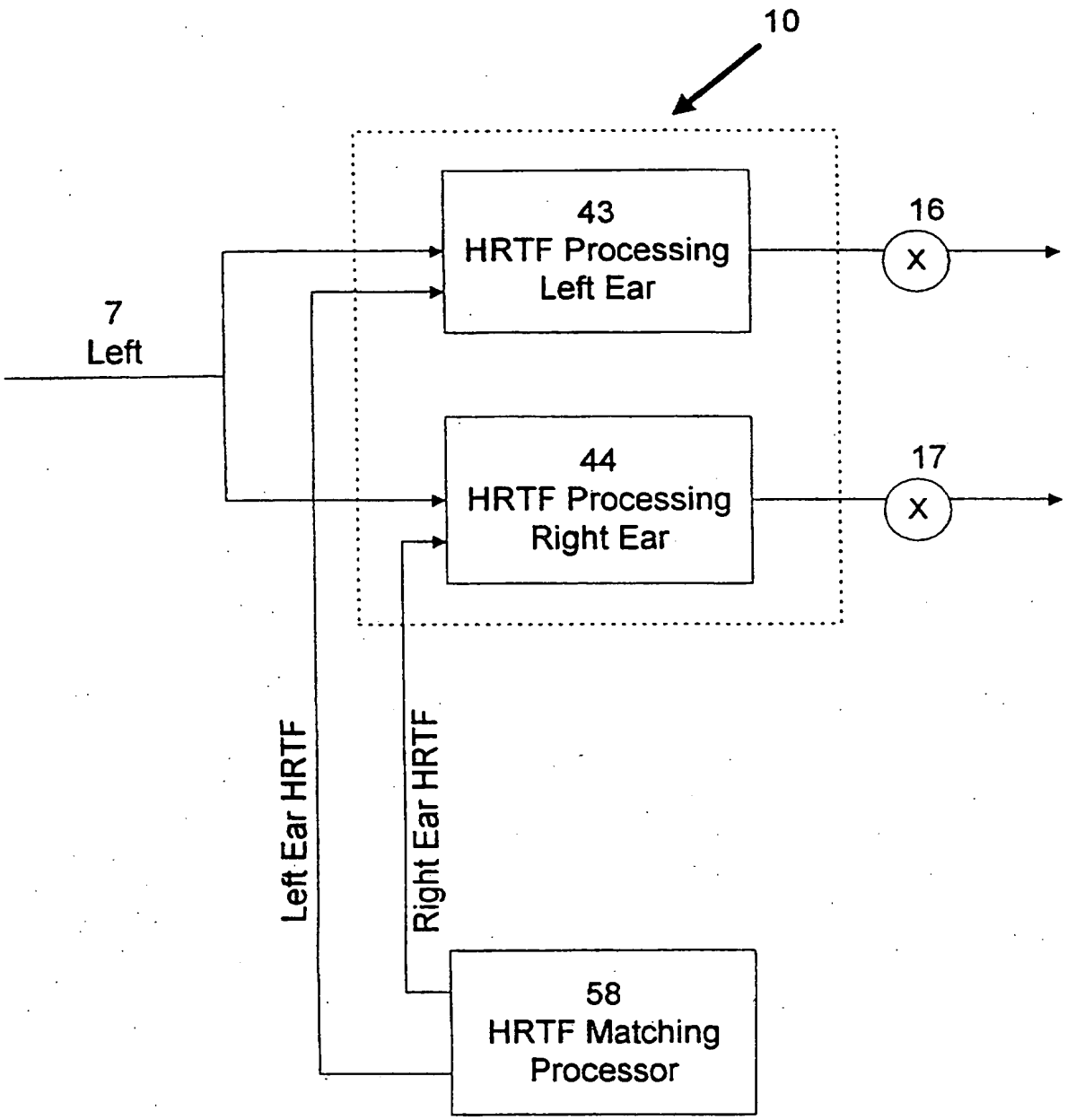


FIG. 6A

NSDOCID: <WO_____9725834A2_I_>

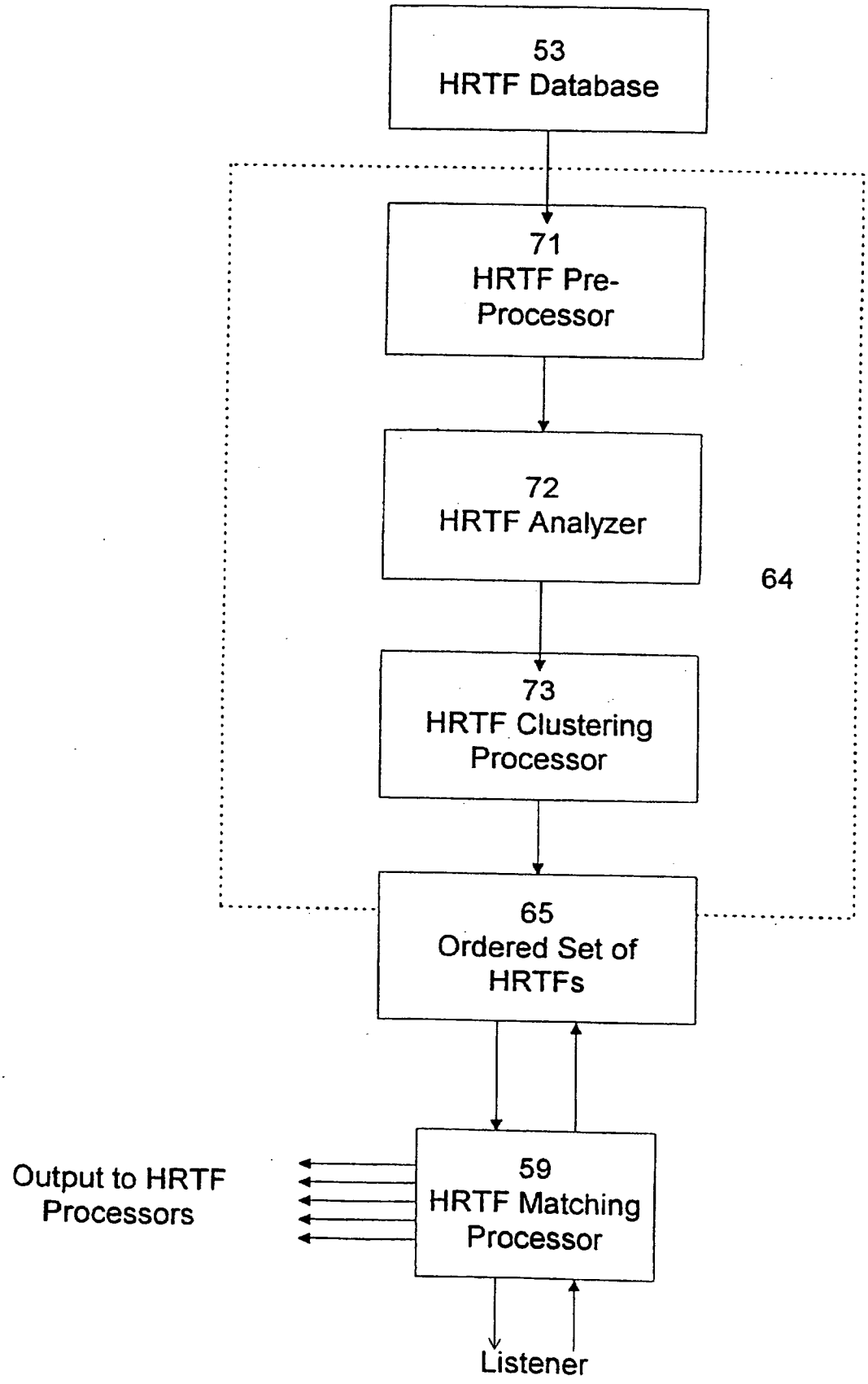

53
HRTF Database
71
HRTF Pre-
Processor
72
HRTF Analyzer
64
73
HRTF Clustering
Processor
65
Ordered Set of
HRTFs
Output to HRTF
Processors
59
HRTF Matching
Processor
Listener

FIG. 6B

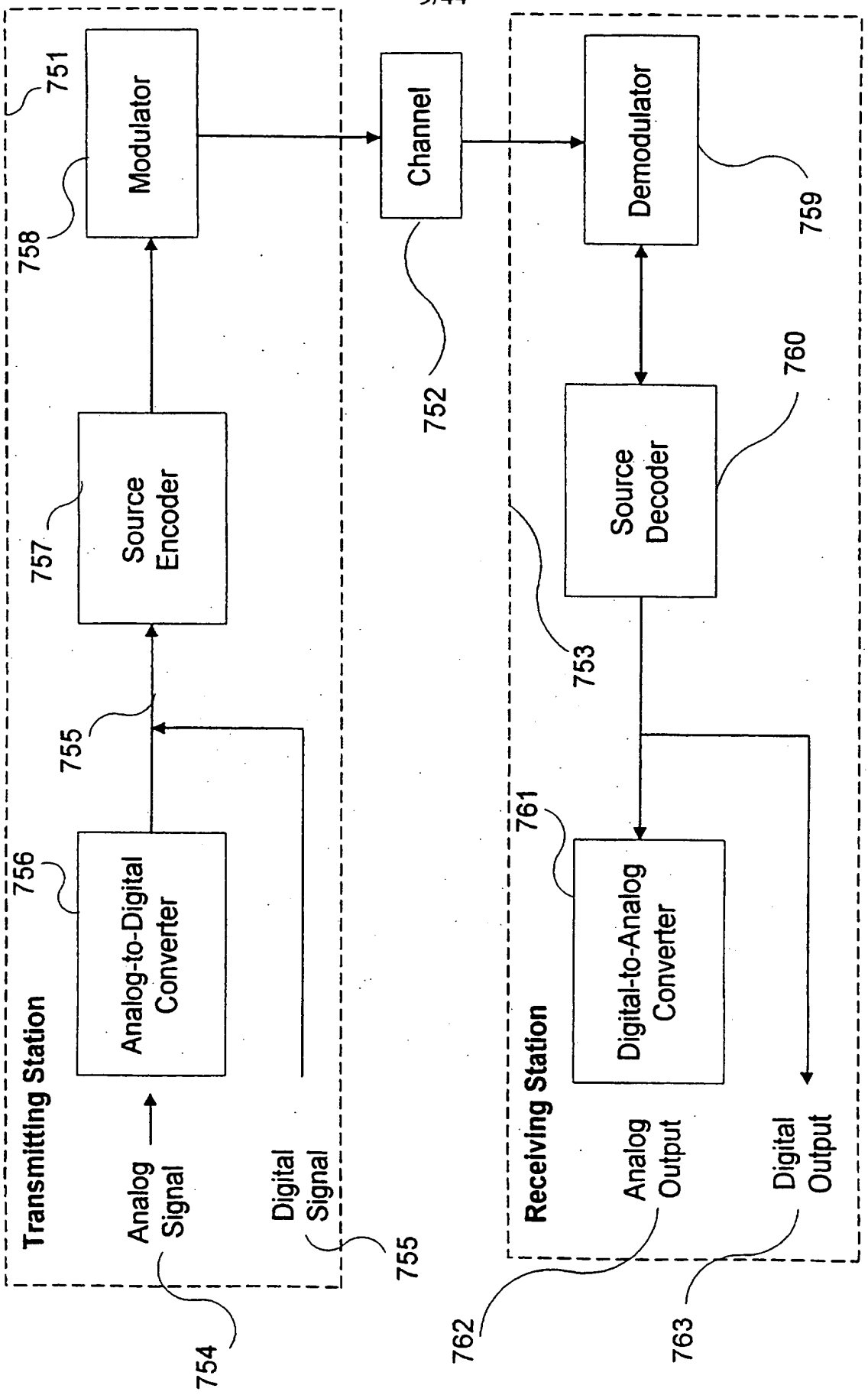


FIG. 7

NSDOCID: <WO_____9725834A2_I_>

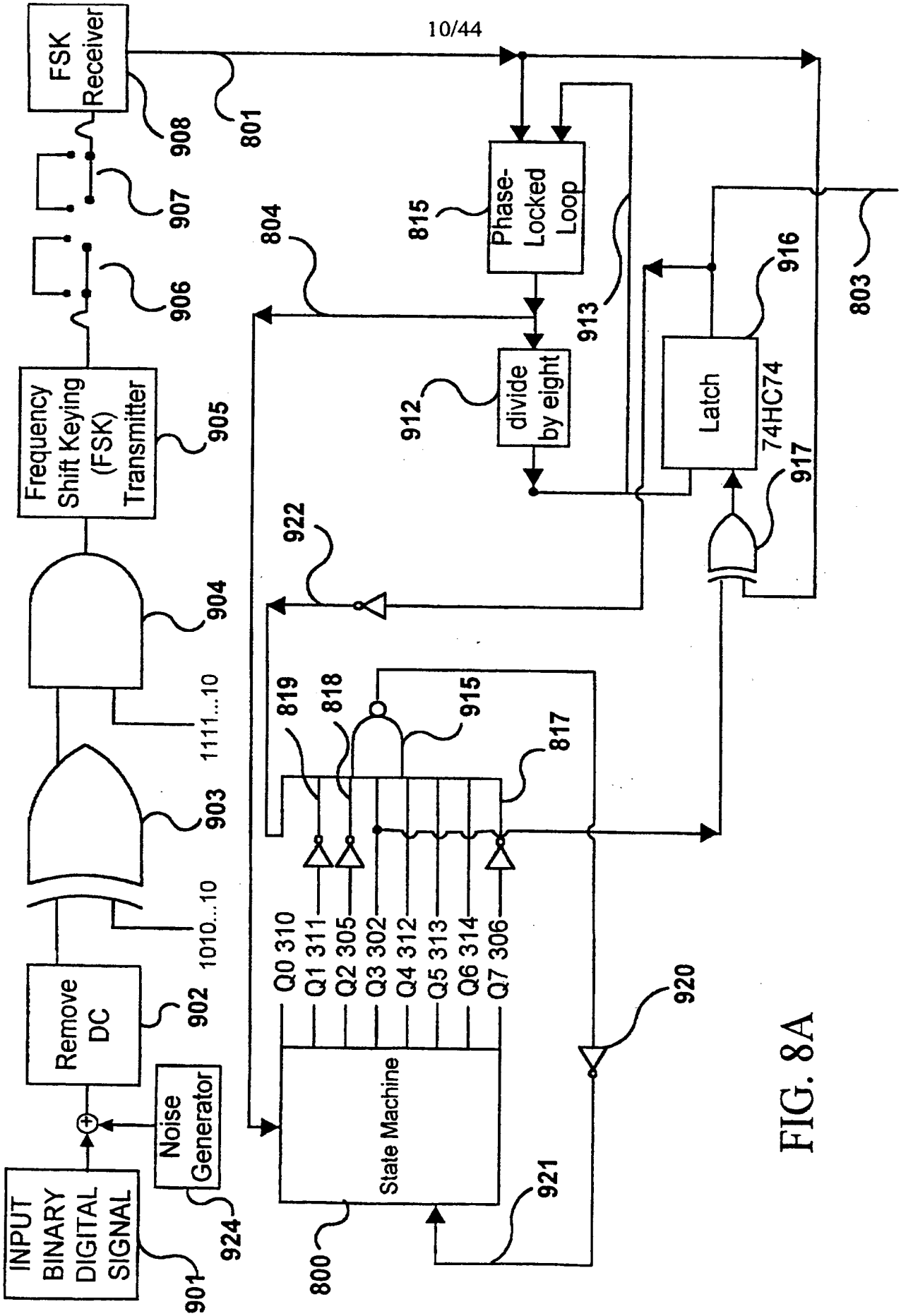


FIG. 8A

SDOCID: <WO_____9725834A2_I_>

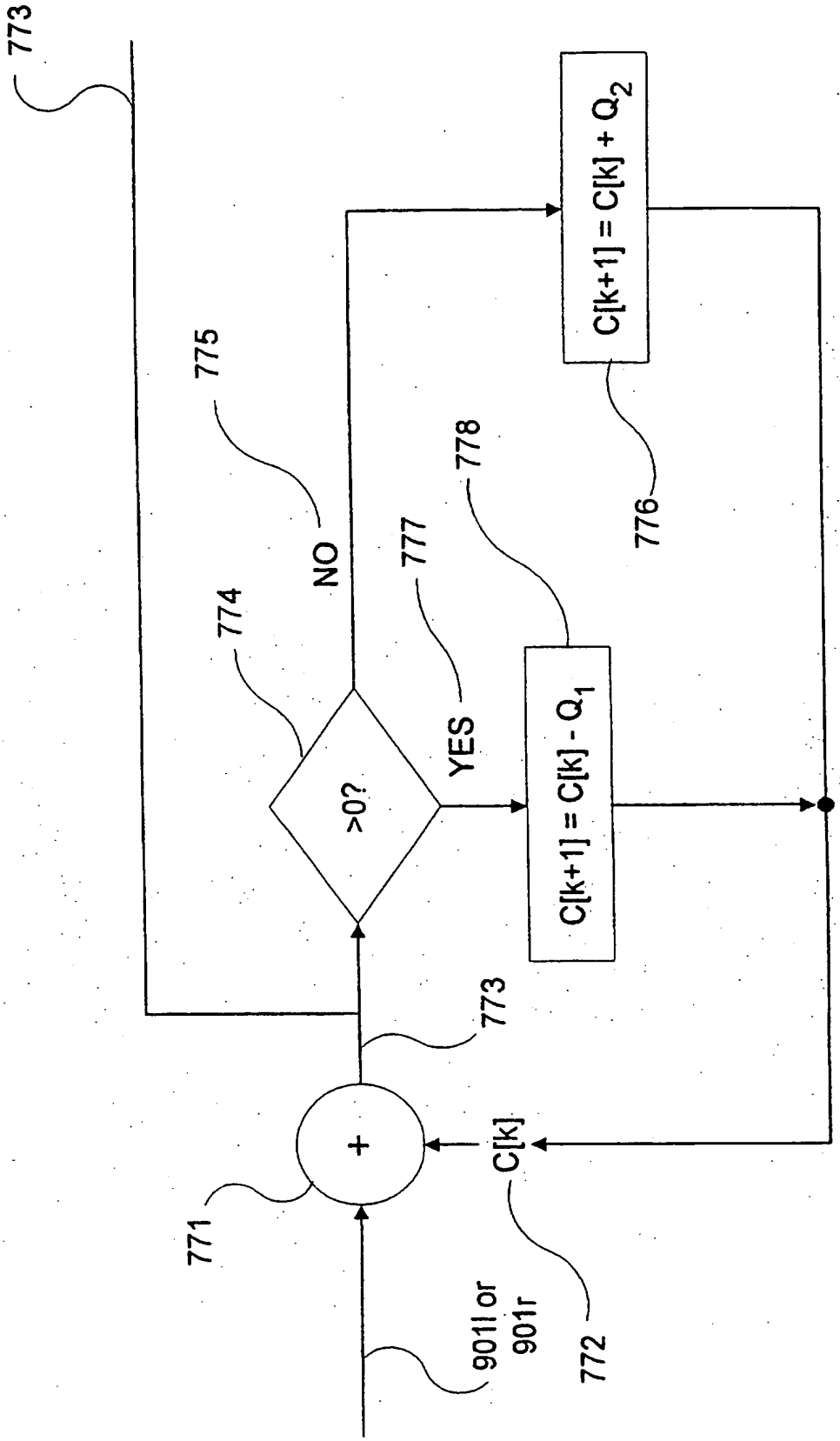

773
775
NO
774
>0?
777
YES
778
C[k+1] = C[k] - Q1
776
C[k+1] = C[k] + Q2
773
771
+
C[k]
772
901l or
901r

FIG. 8B

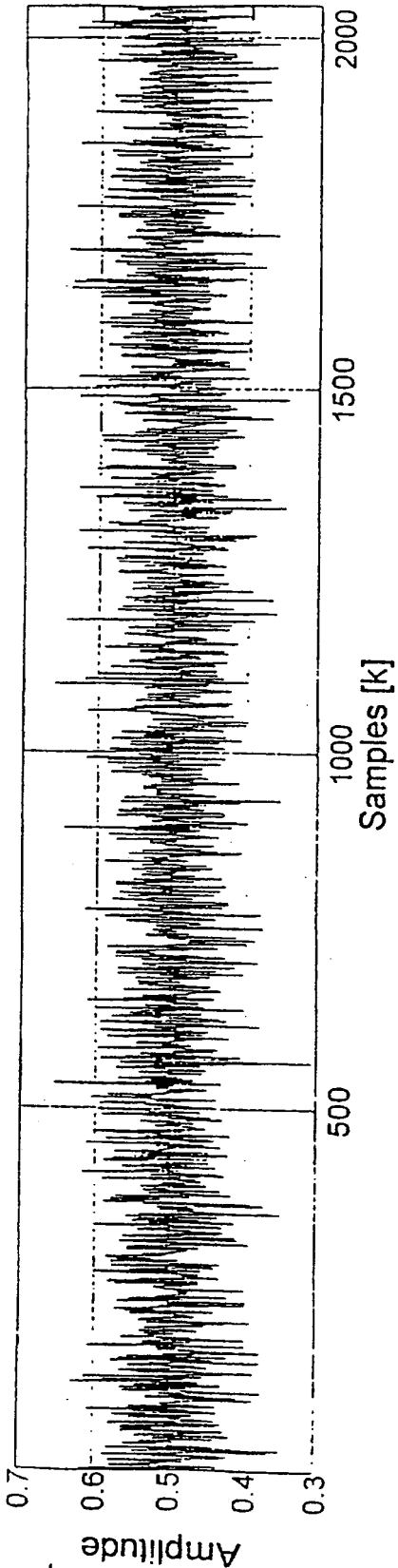


FIG. 9A

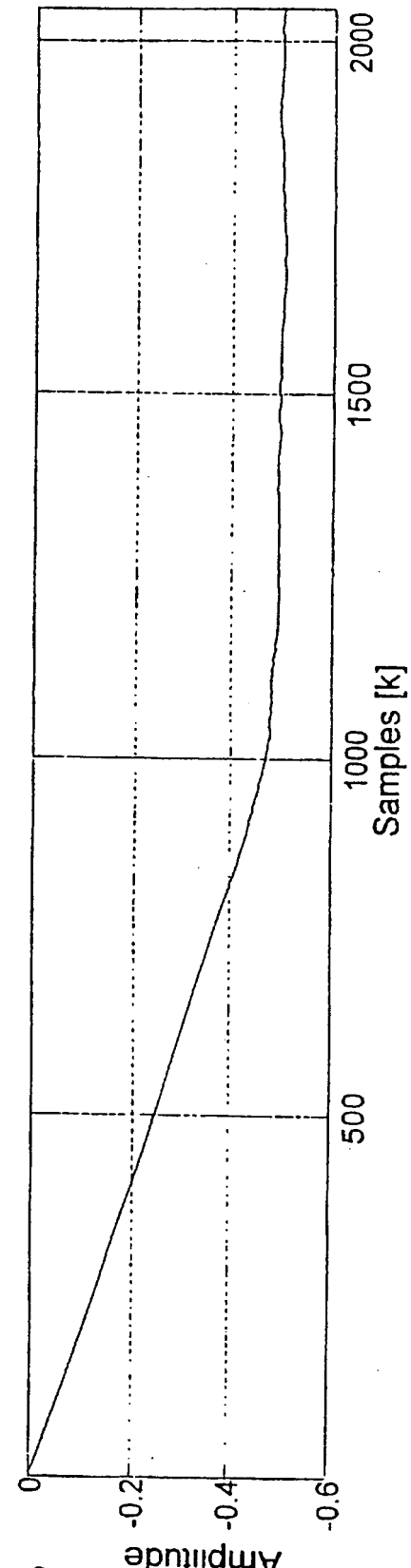


FIG. 9B

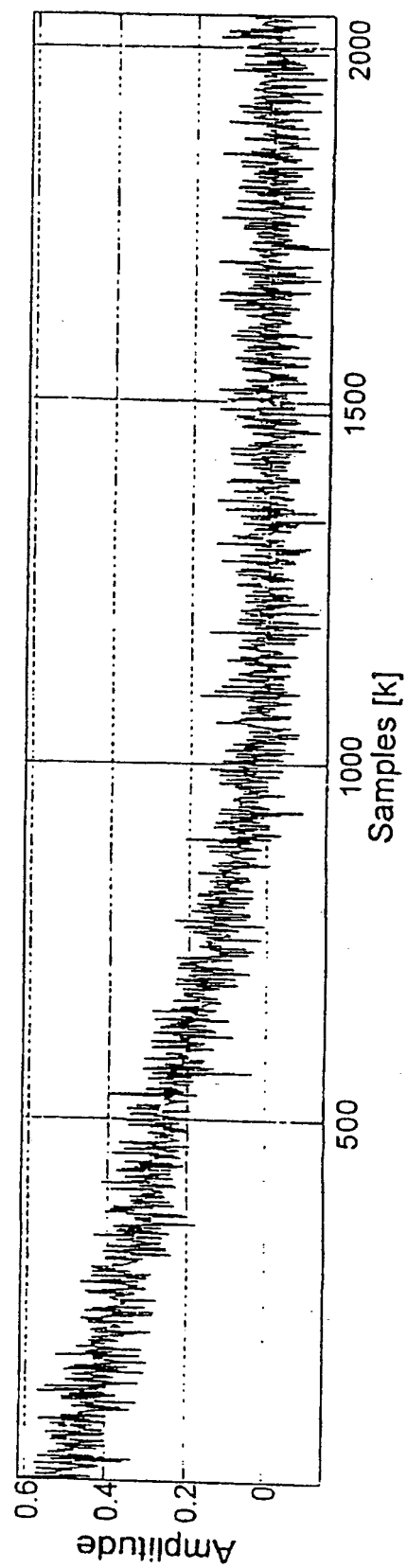


FIG. 9C

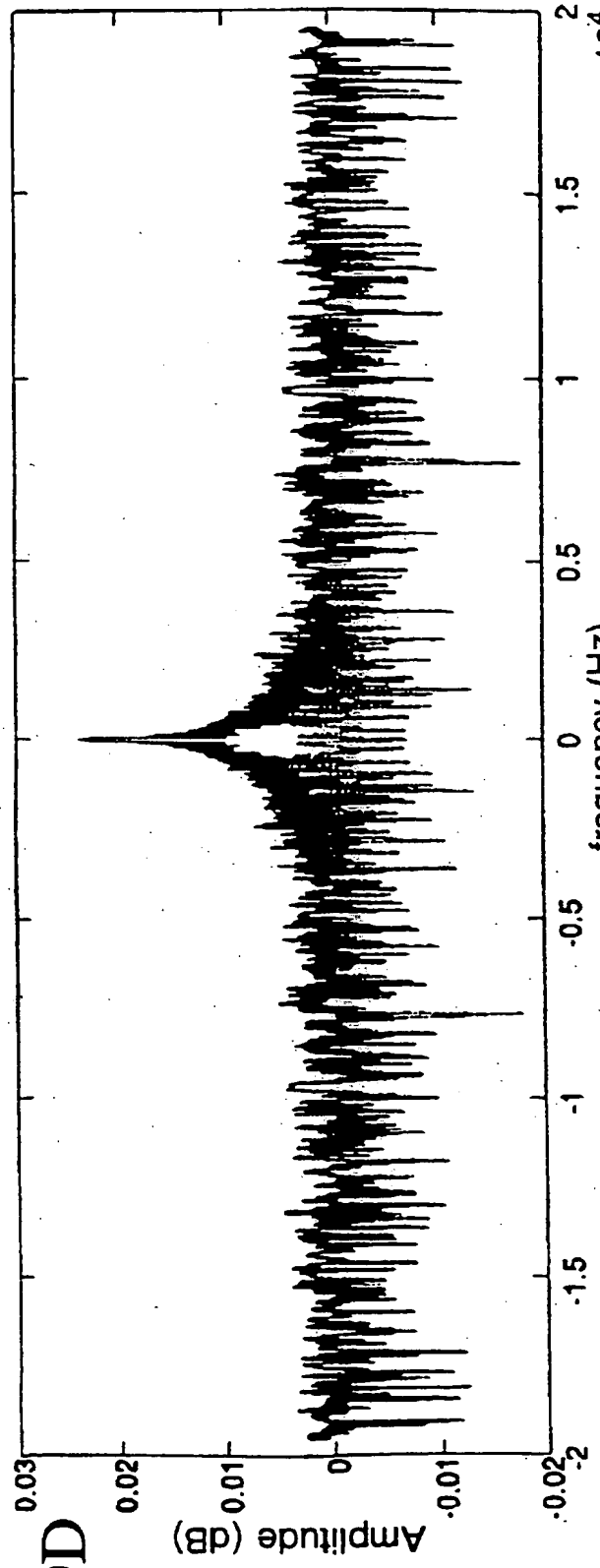


FIG. 9D

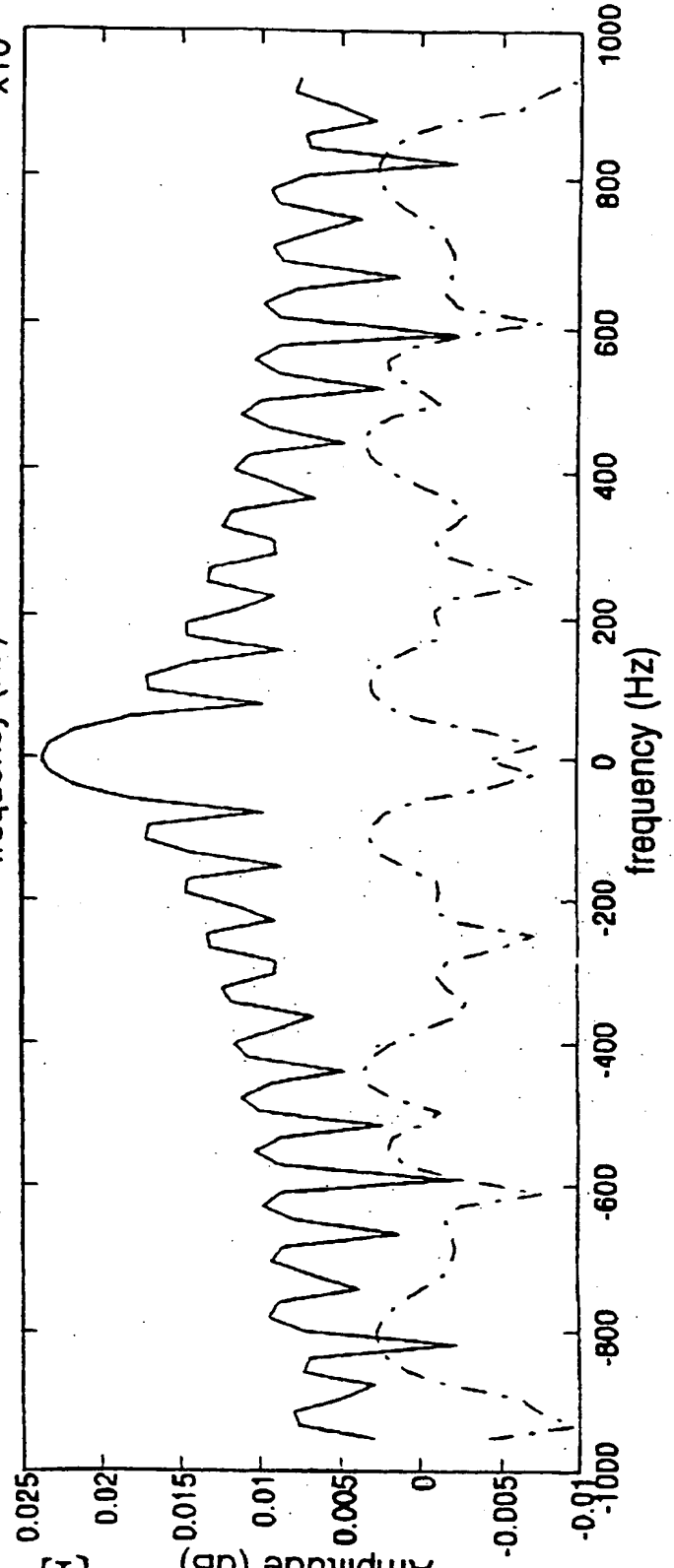


FIG. 9E

NSDOCID: <WO_____9725834A2_I_>

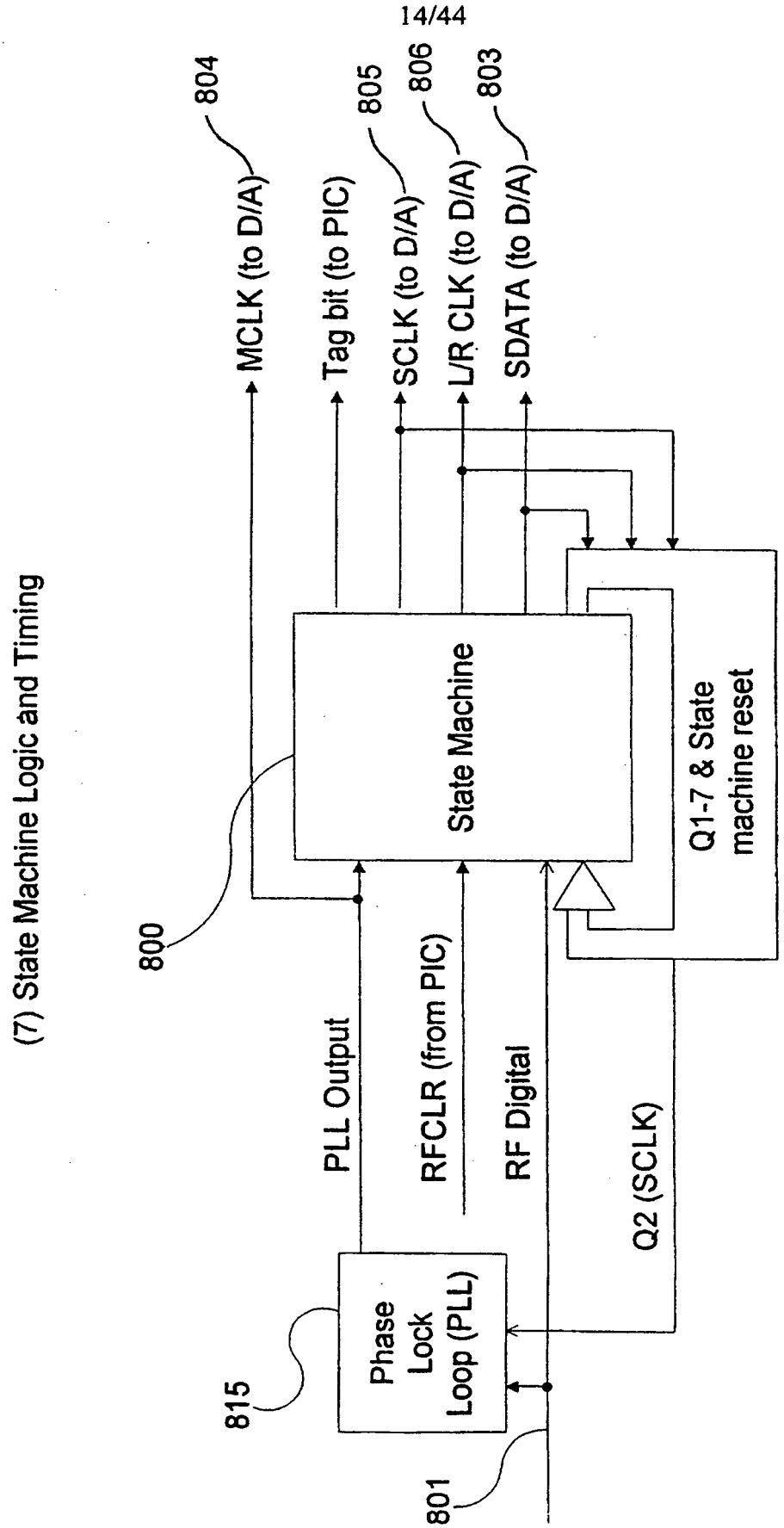

(7) State Machine Logic and Timing
804
MCLK (to D/A)
805
Tag bit (to PIC)
806
SCLK (to D/A)
803
L/R CLK (to D/A)
SDATA (to D/A)
State Machine
Q1-7 & State machine reset
800
PLL Output
RFCLR (from PIC)
RF Digital
Q2 (SCLK)
Phase Lock Loop (PLL)
815
801

FIG. 9F

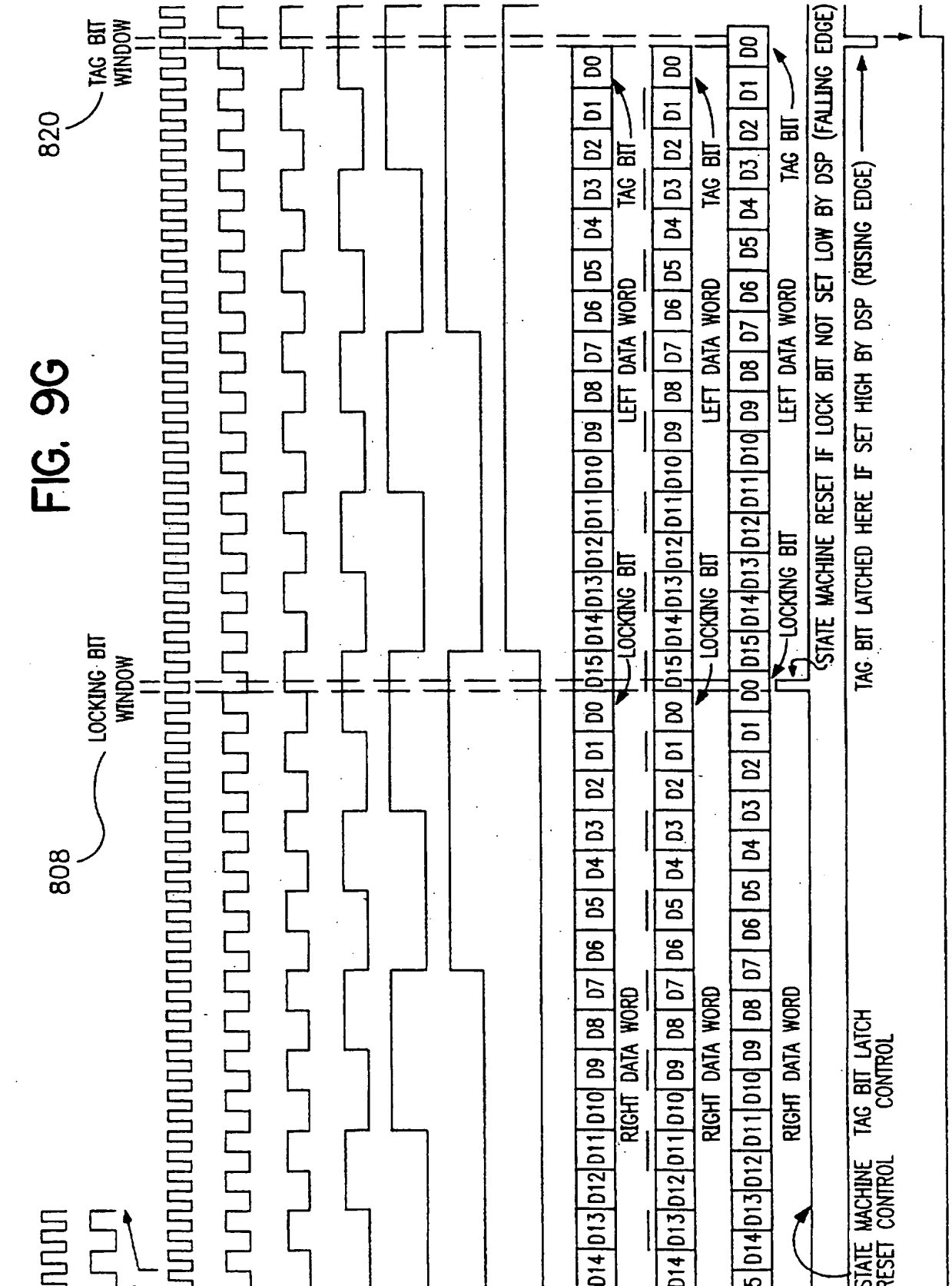


FIG. 9G

INSDOCID: <WO_____9725834A2_I_>

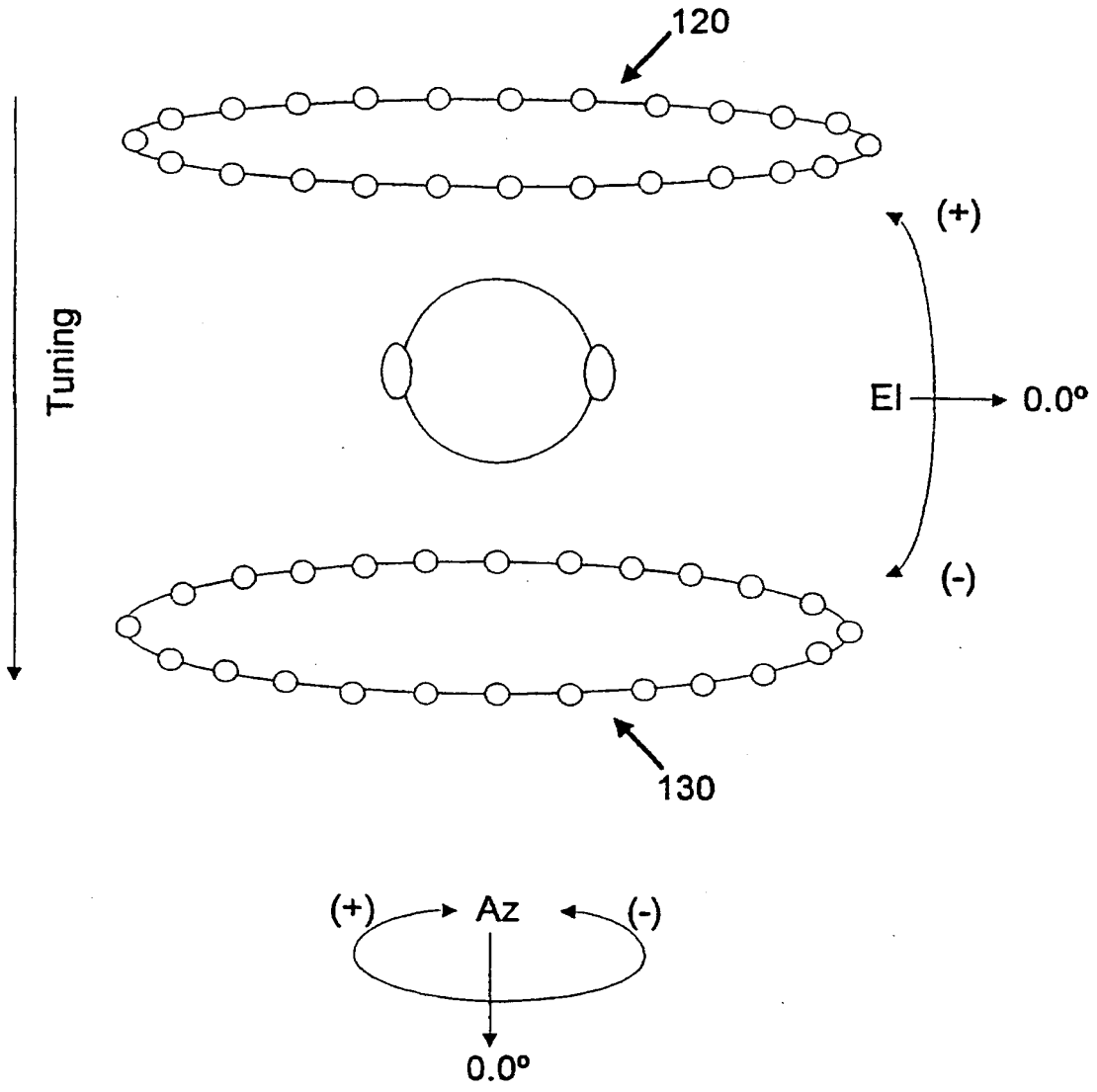


FIG. 10

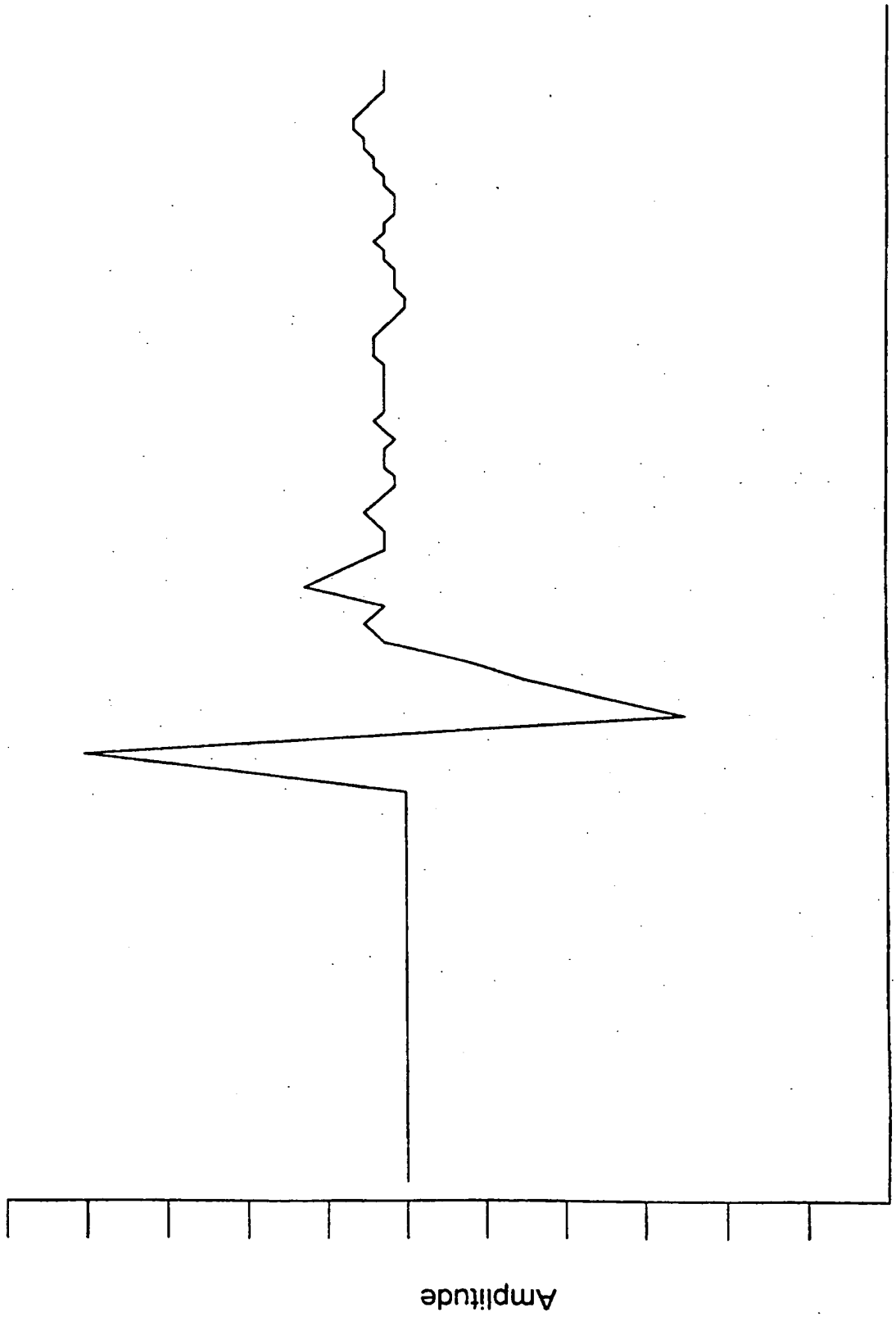


FIG. 11

NSDOCID: <WO_____9725834A2_I_>

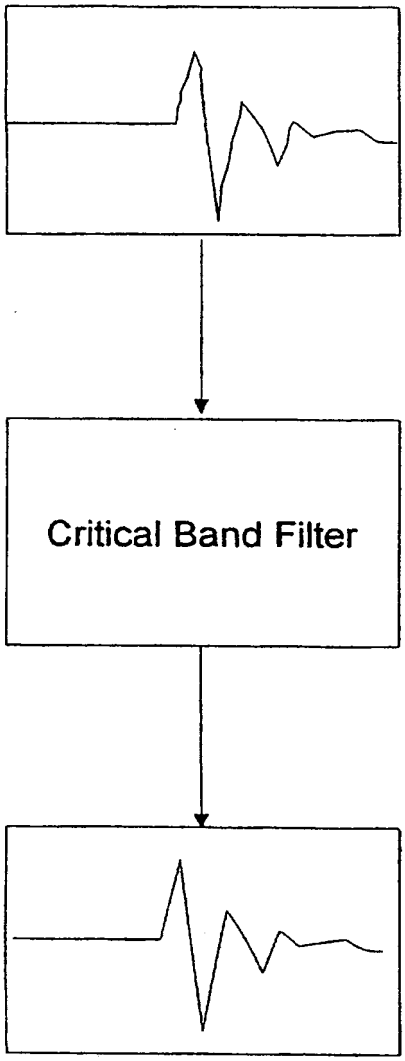


FIG 12

JSDOCID: <WO_____9725834A2_I_>

	HRTF Location 1				HRTF Location 2				...	HRTF Location L			
Subject	Band 1	Band 2	...	Band N	Band 1	Band 2	...	Band N	...	Band 1	Band 2	Band N	
1 right													
1 left													
2 right													
2 left													
...													
S right													
S left													

FIG. 13

NSDOCID: <WO_____9725834A2_I_>

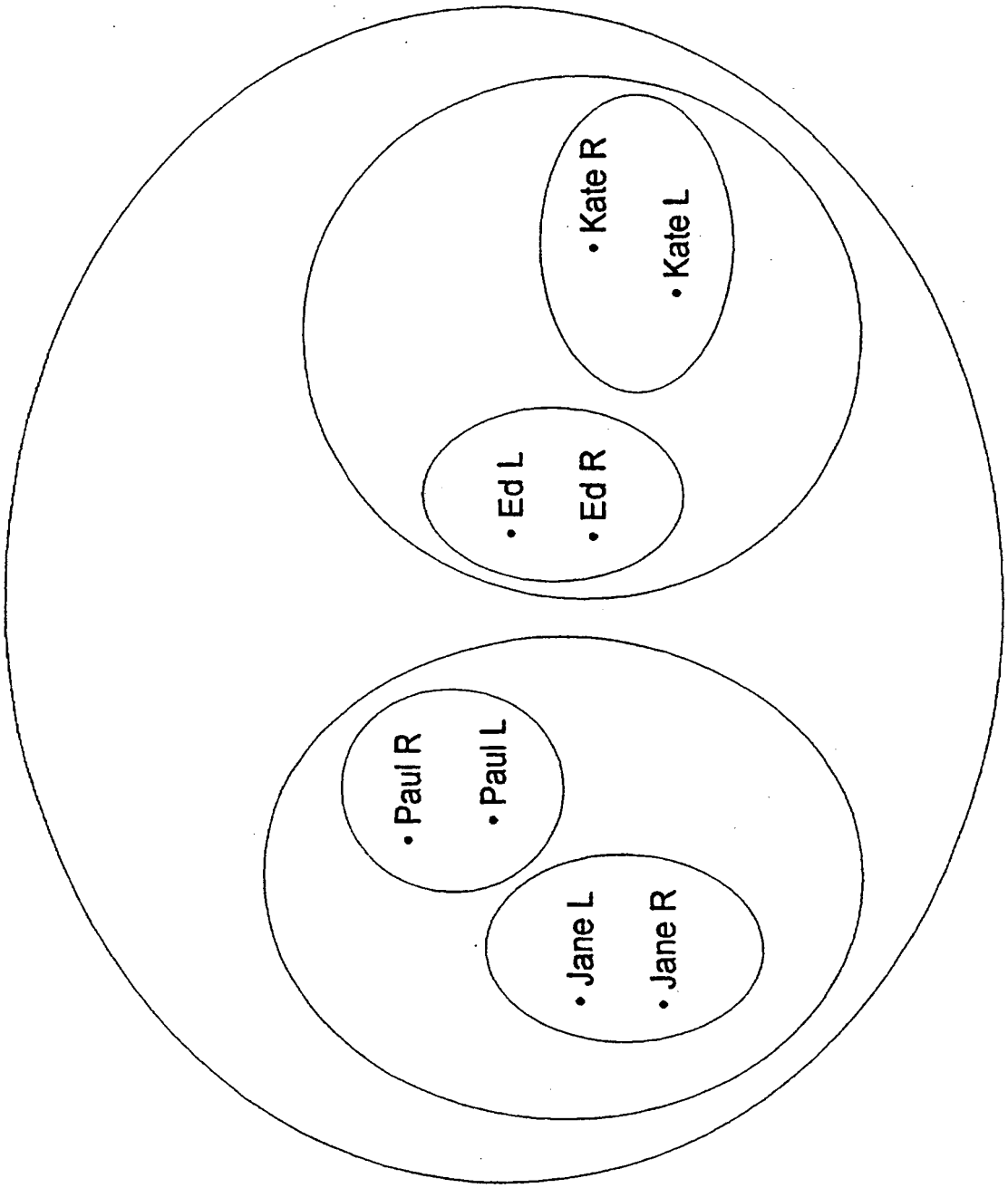


FIG. 14

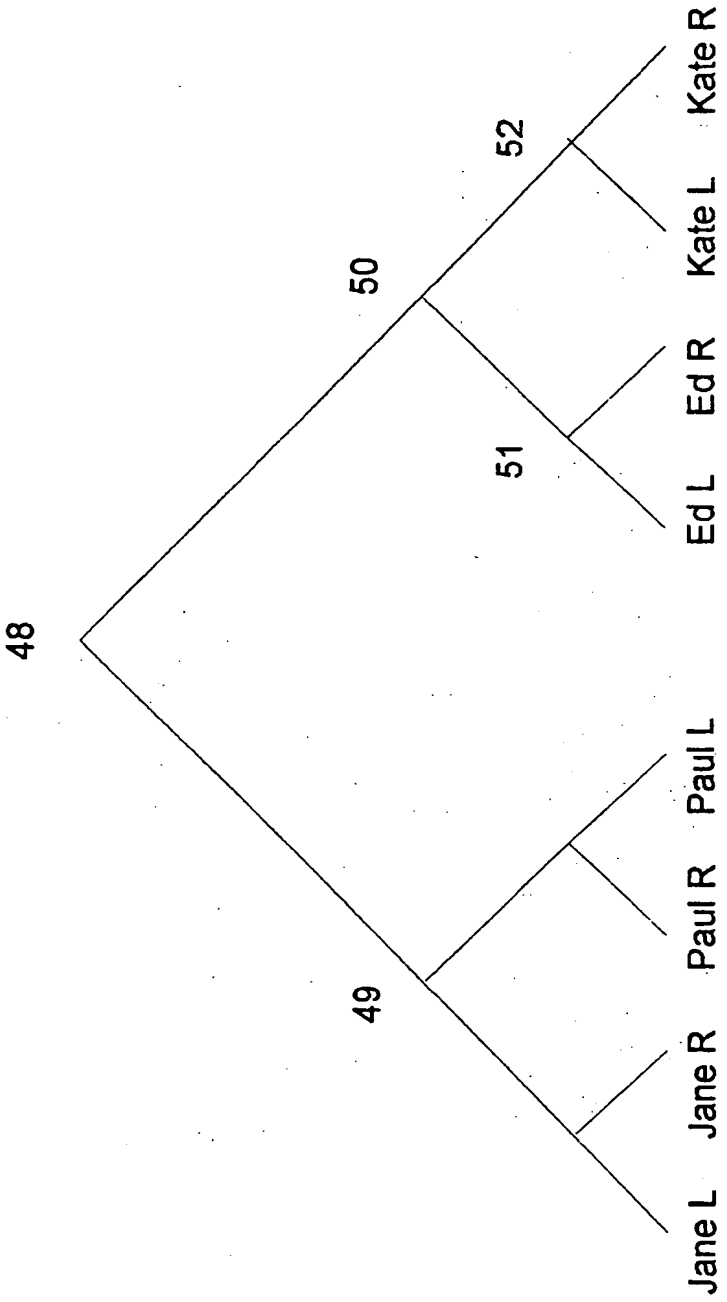


FIG. 15

NSDOCID: <WO_____9725834A2_I_>

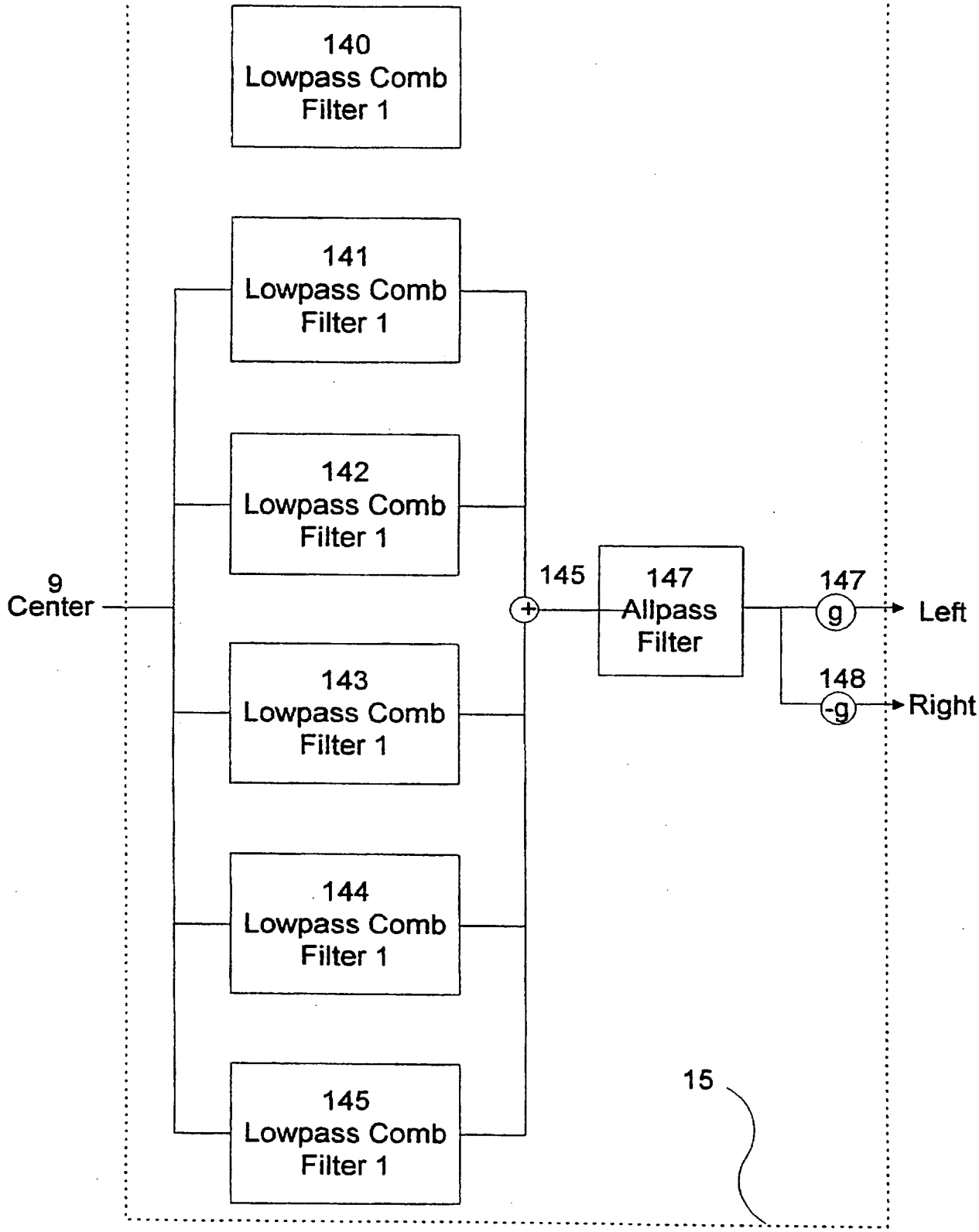

140
Lowpass Comb
Filter 1
141
Lowpass Comb
Filter 1
142
Lowpass Comb
Filter 1
143
Lowpass Comb
Filter 1
144
Lowpass Comb
Filter 1
145
Lowpass Comb
Filter 1
9
Center
145
147
Allpass
Filter
147
g
Left
148
-g
Right
15

FIG. 16

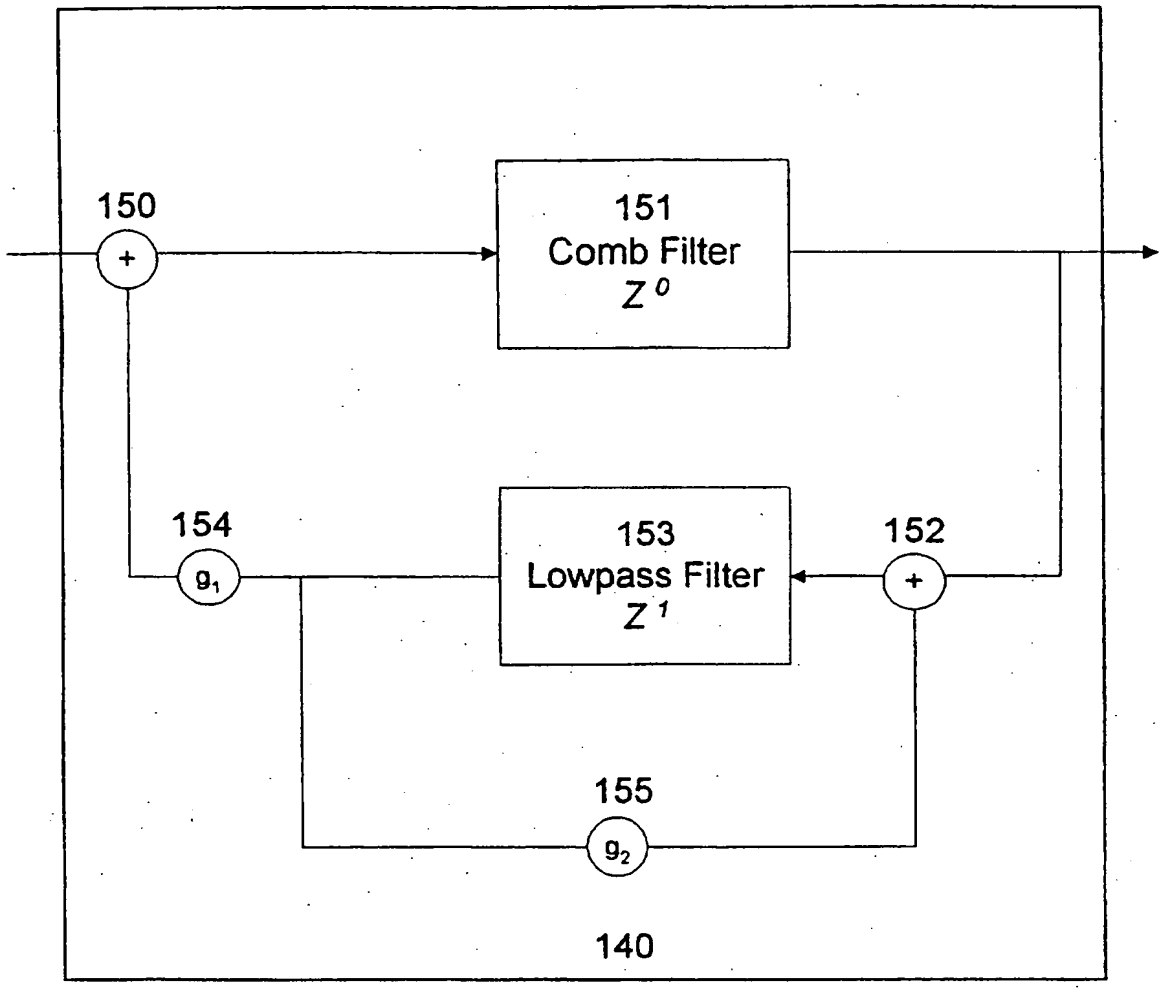


FIG. 17

INSDOCID: <WO_____9725834A2_I_>

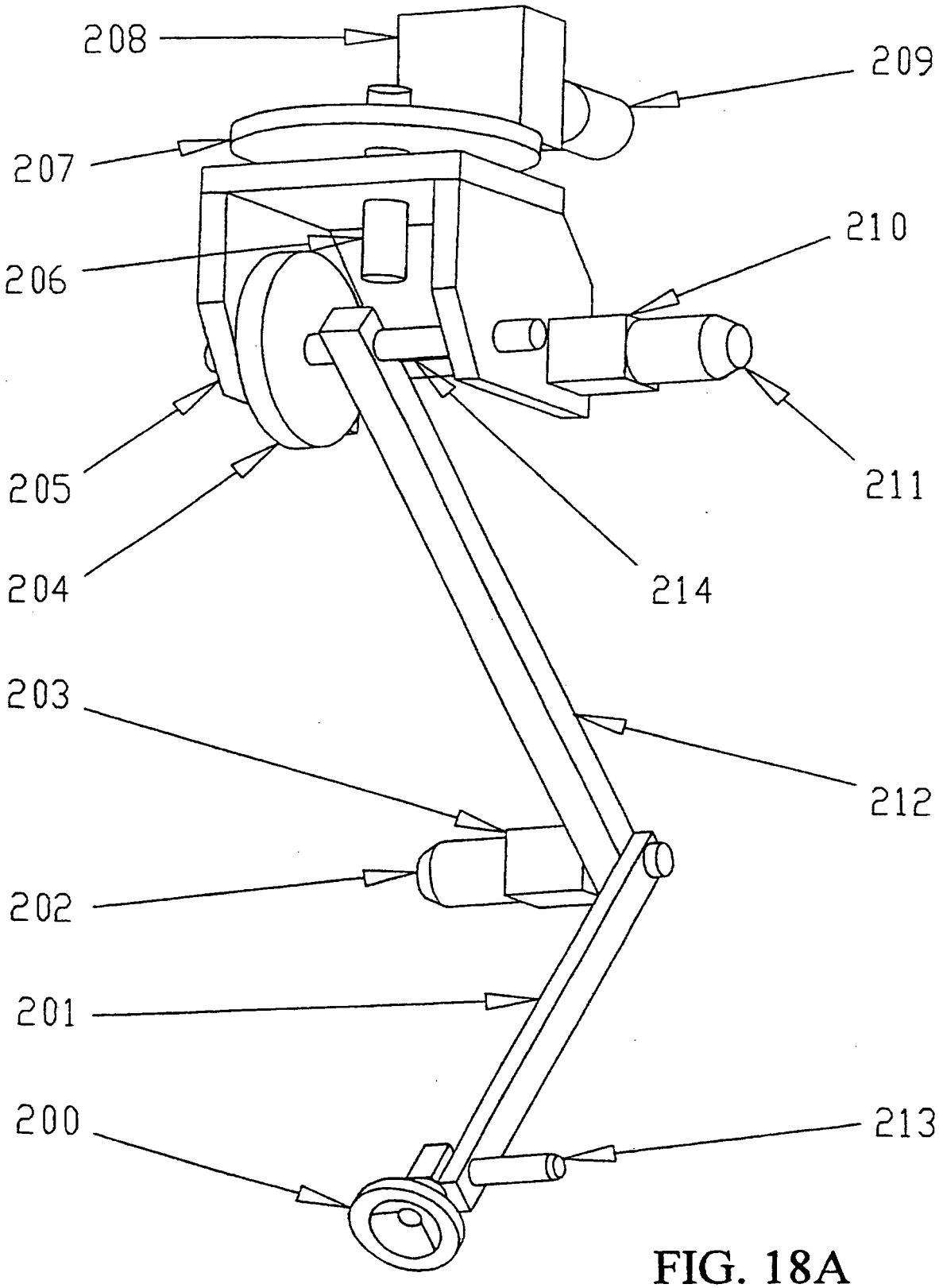


FIG. 18A

SDOCID: <WO_____9725834A2_I_>

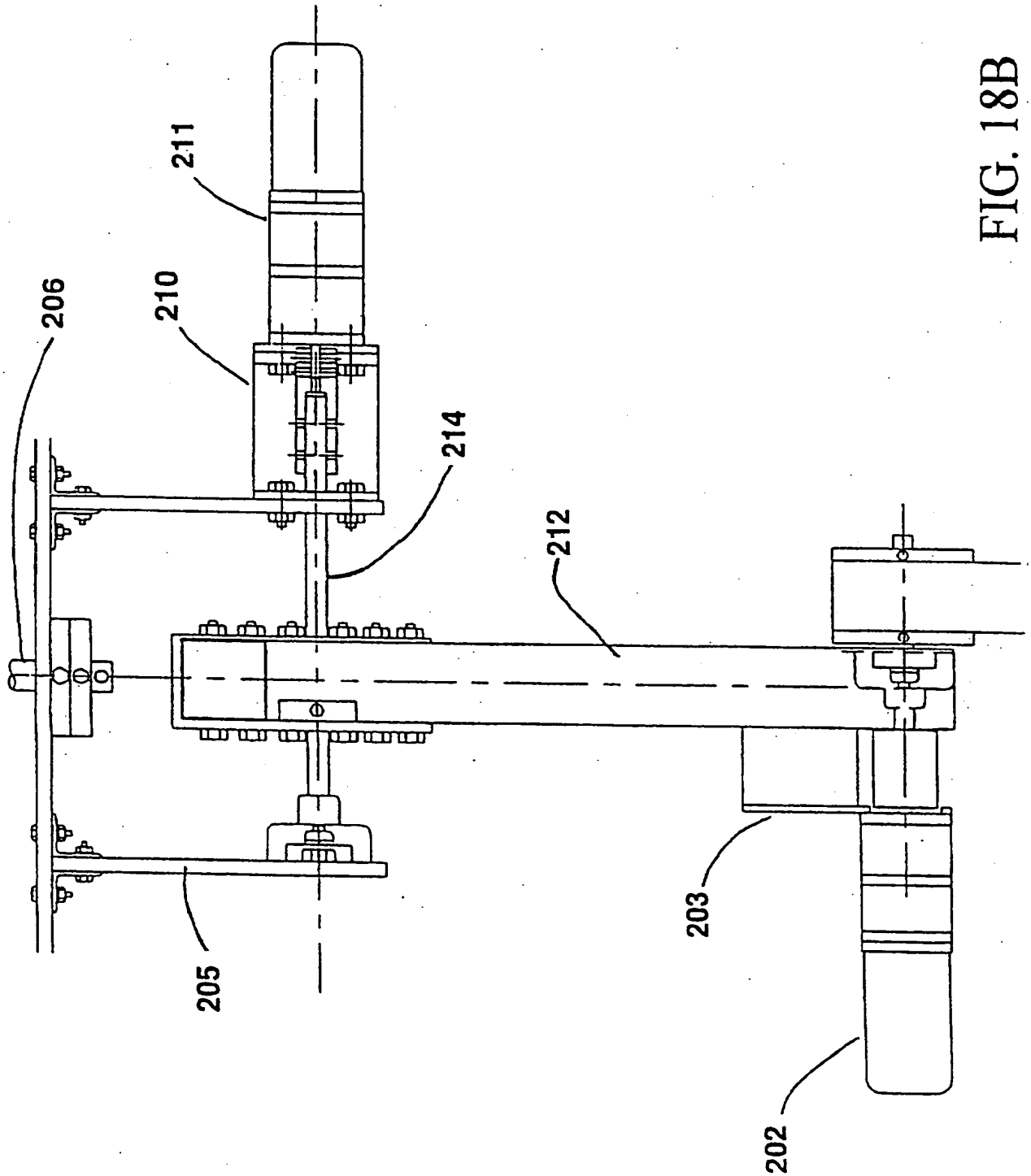


FIG. 18B

NSDOCID: <WO_____9725834A2_I_>

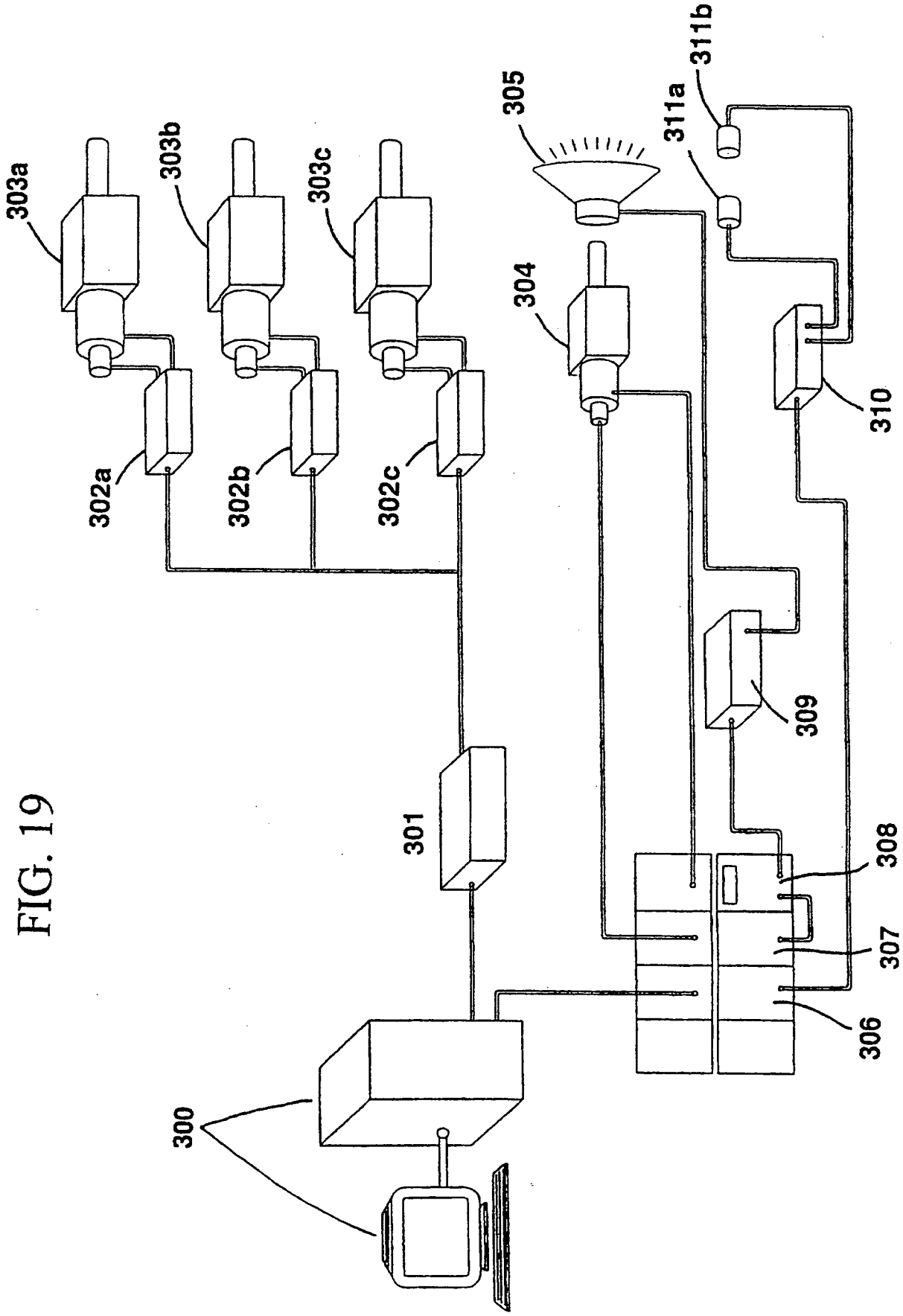

303a
303b
303c
305
311b
311a
304
310
302a
302b
302c
309
301
308
307
306
300

FIG. 19

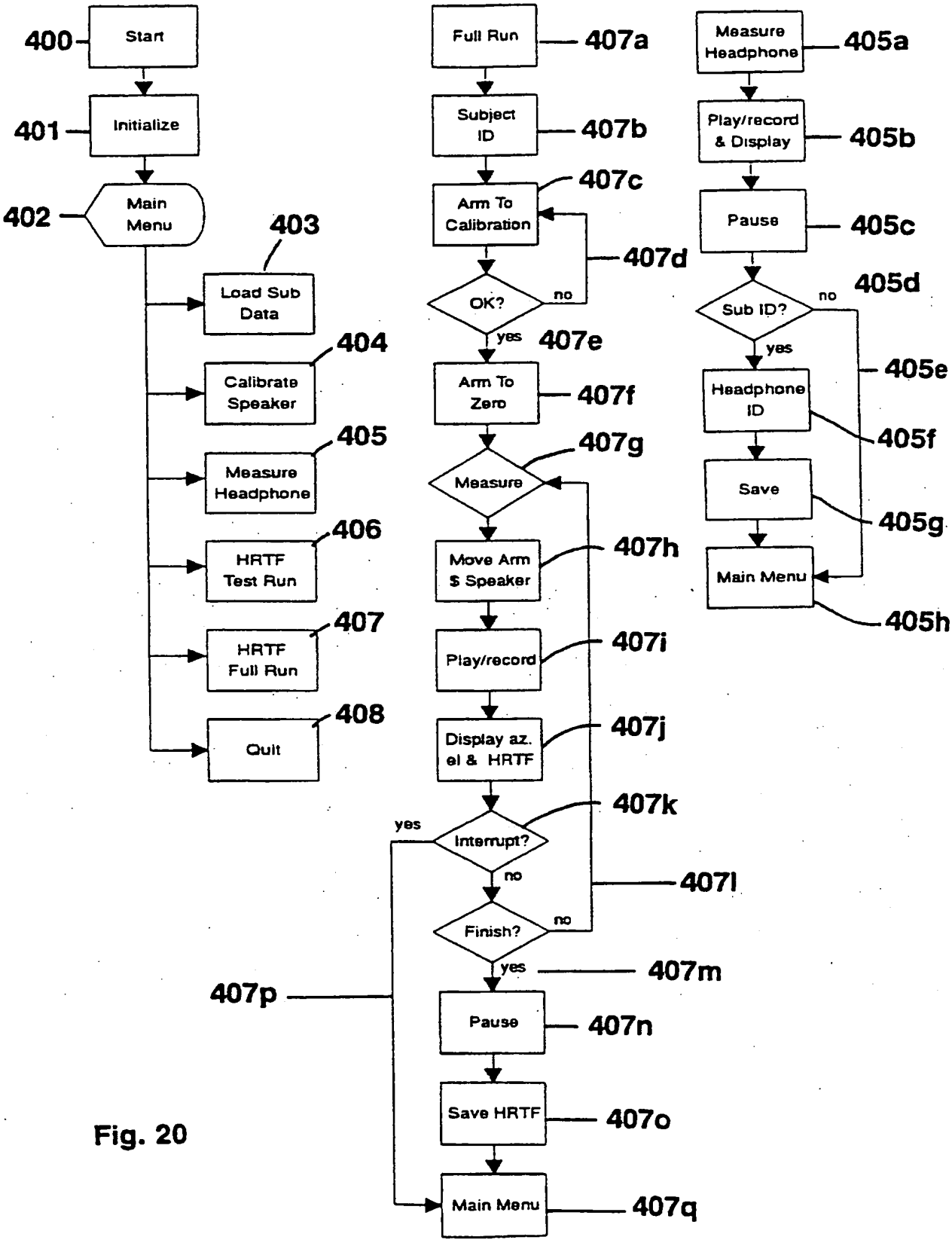


Fig. 20

INSDOCID: <WO_____9725834A2_I_>

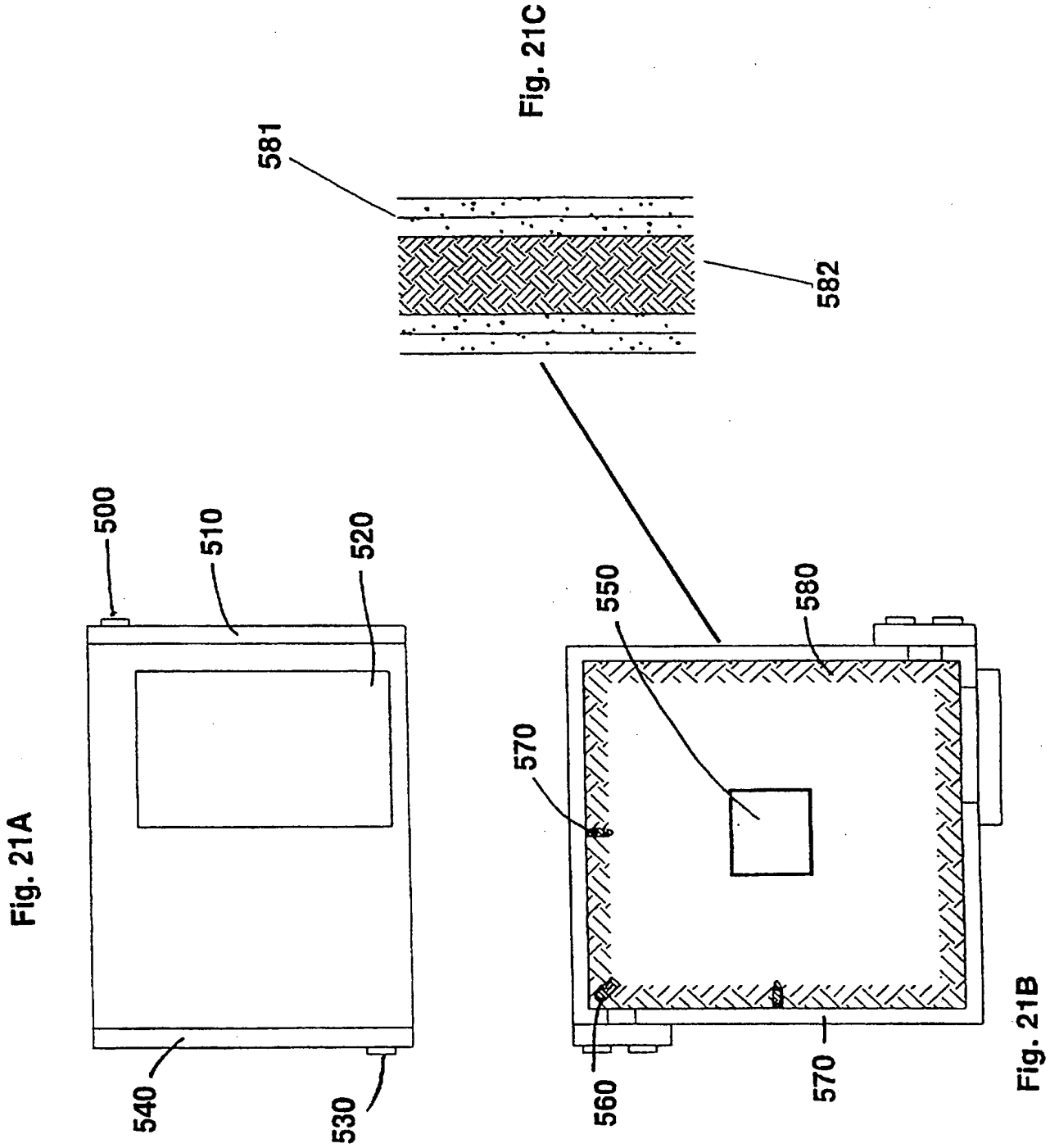


Fig. 21C

Fig. 21A

Fig. 21B

ISDOCID: <WO_____9725834A2_I_>

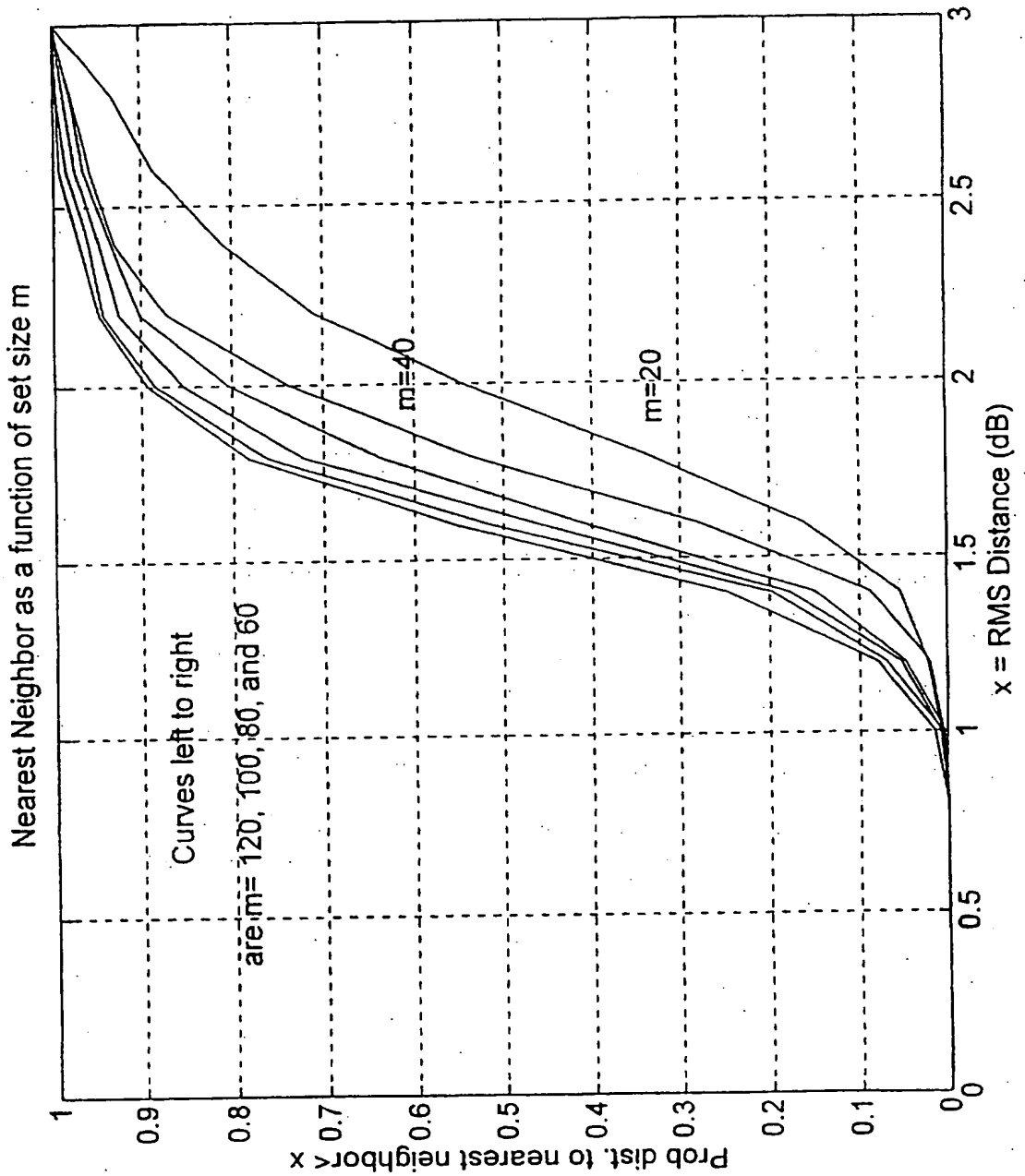
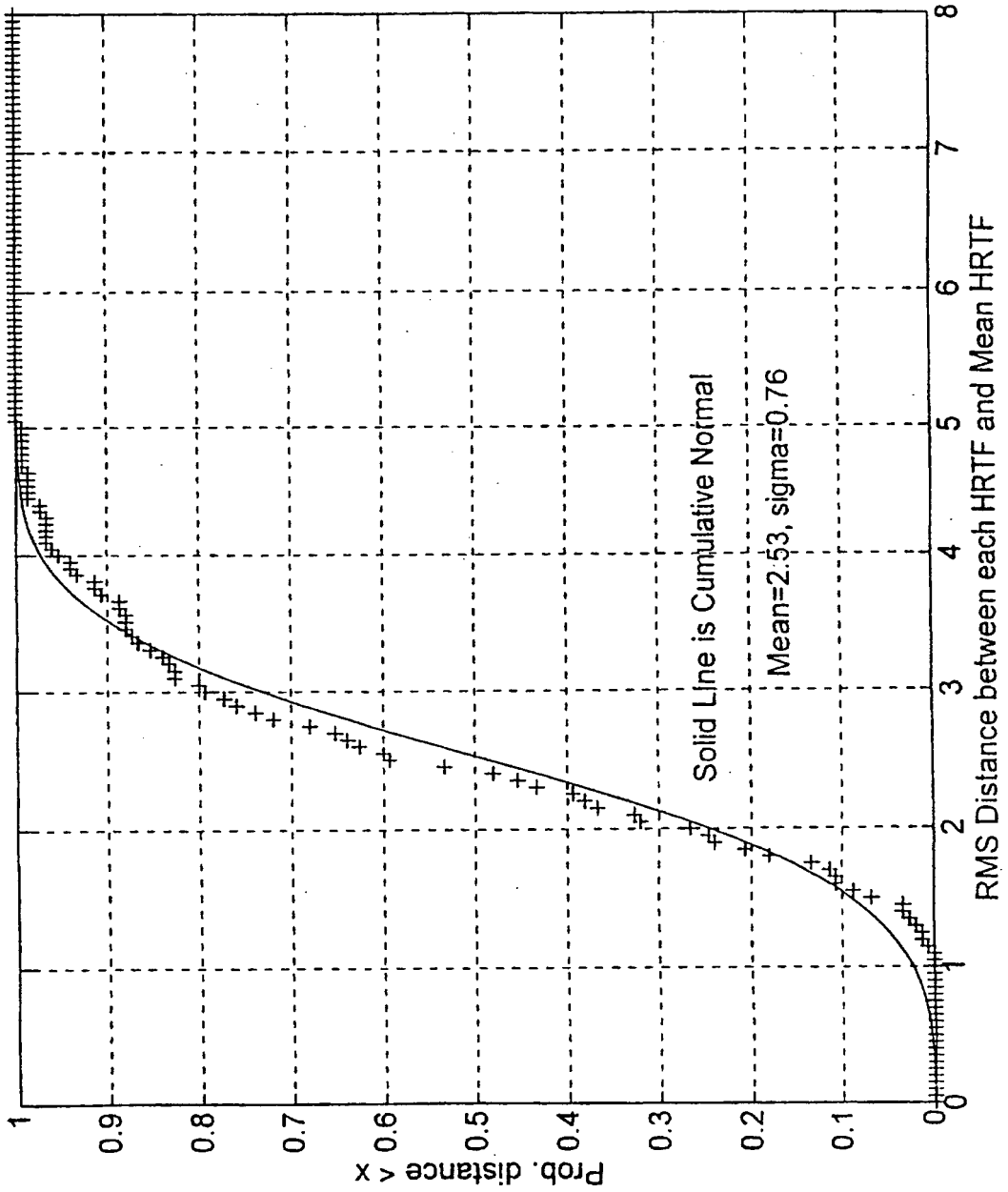


Fig. 22A

NSDOCID: <WO_____9725834A2_I_>

FIG. 22B


Prob. distance < x
RMS Distance between each HRTF and Mean HRTF
Solid Line is Cumulative Normal
Mean=2.53, sigma=0.76
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
1
2
3
4
5
6
7
8

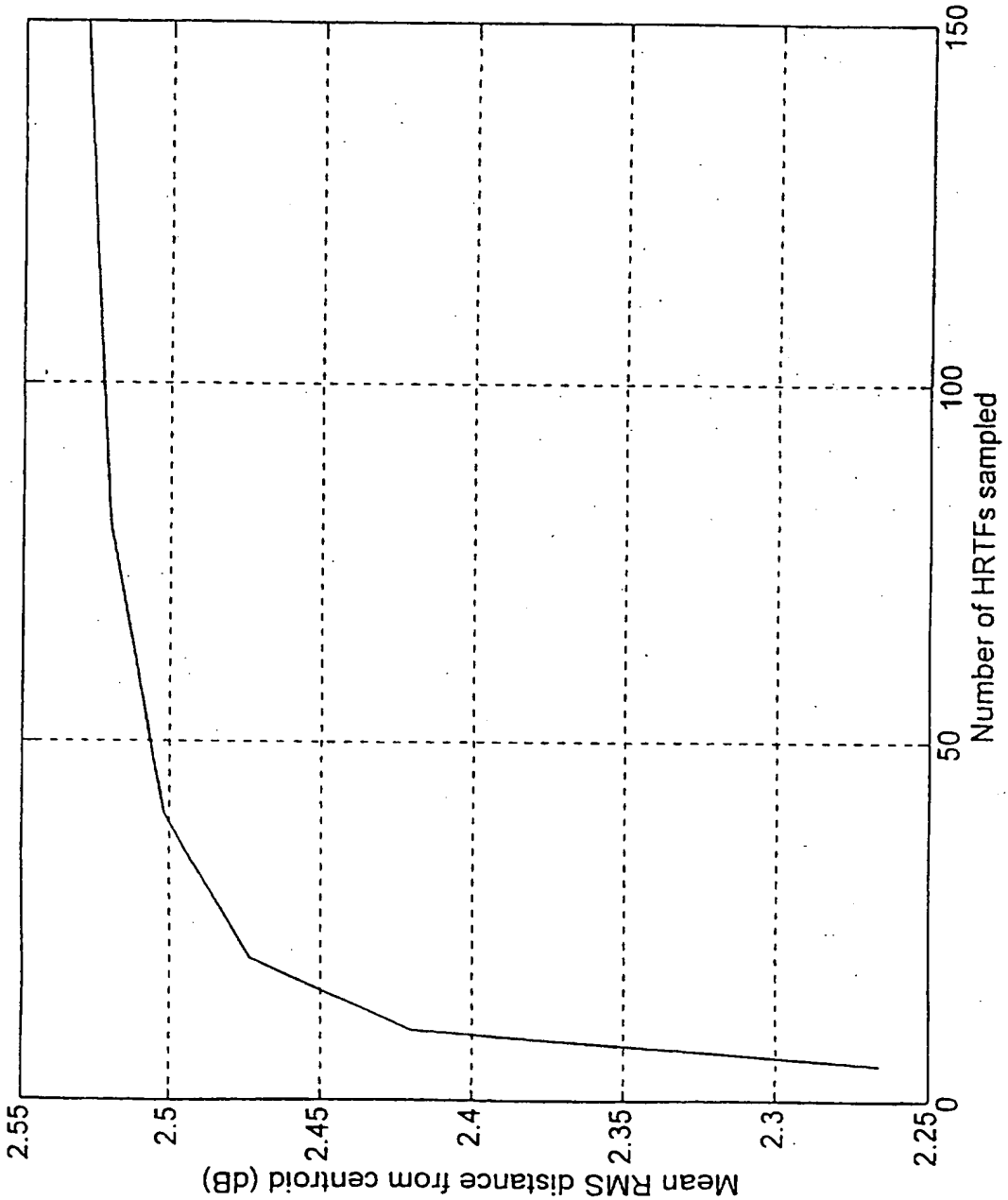


FIG. 22C

ENSDOCID: <WO_____9725834A2_I_>

FIG. 22D

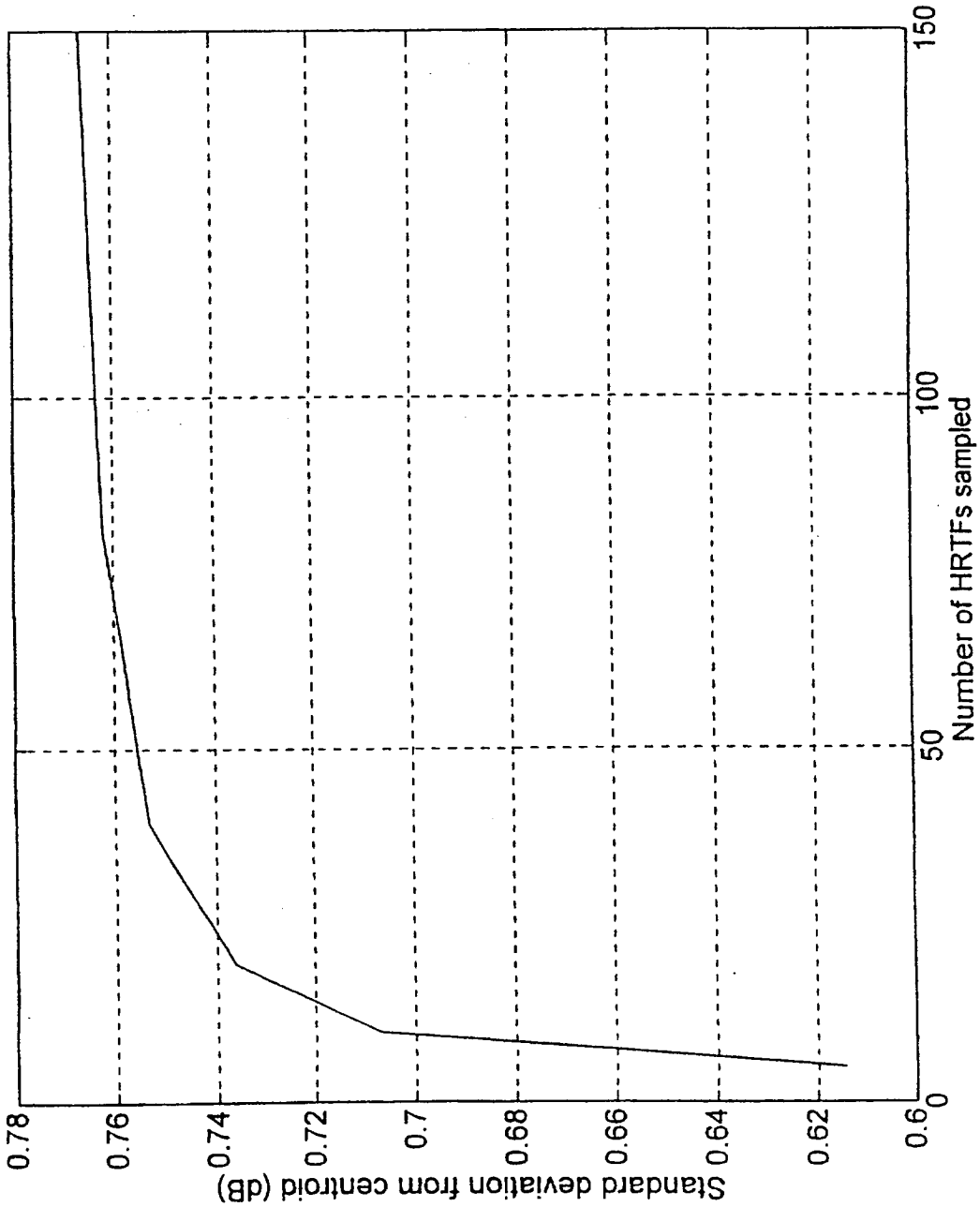
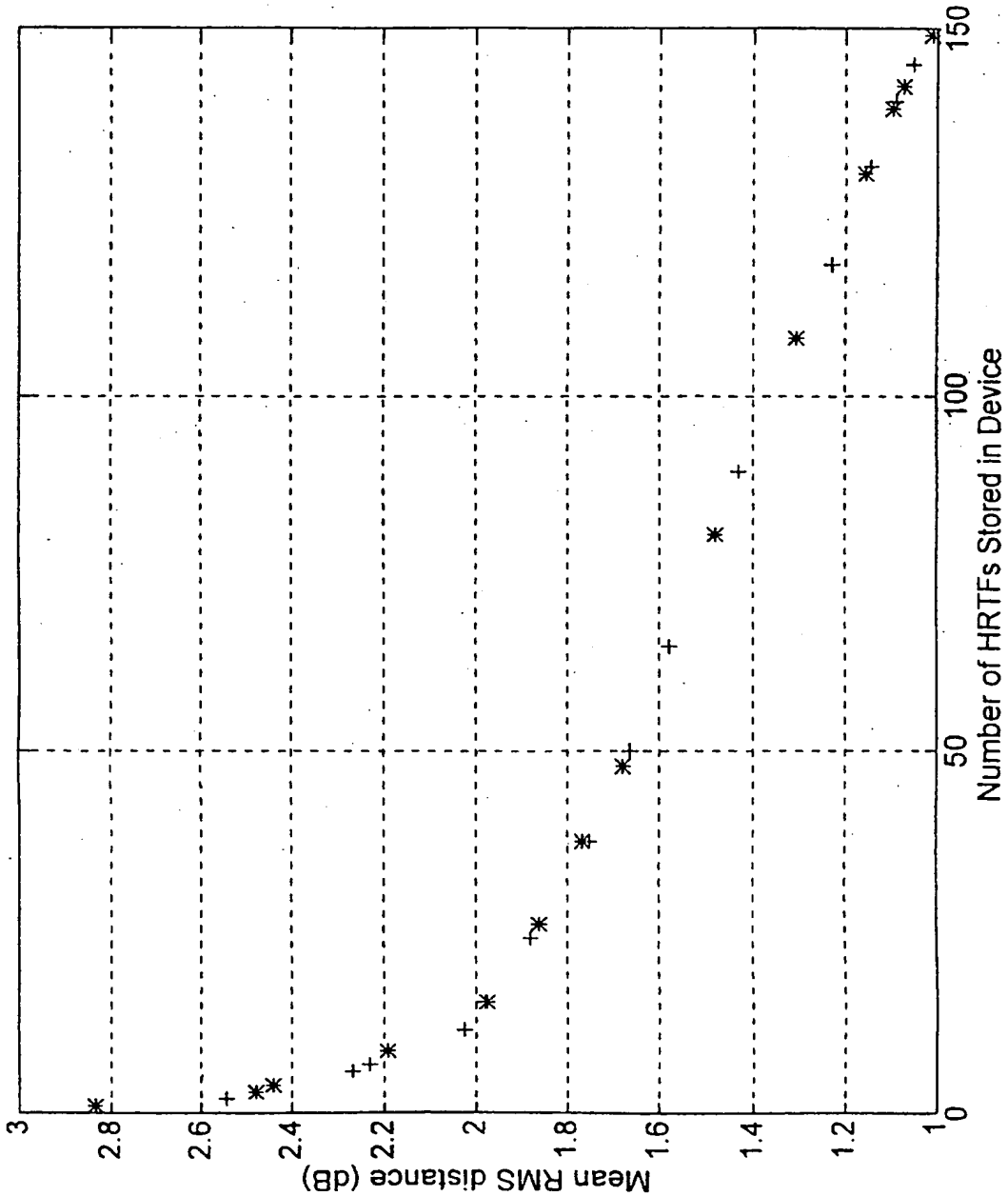

Standard deviation from centroid (dB)
0.78
0.76
0.74
0.72
0.7
0.68
0.66
0.64
0.62
0.6
0
50
100
150
Number of HRTFs sampled

FIG. 22E



INSDOCID: <WO_____9725834A2_I_>

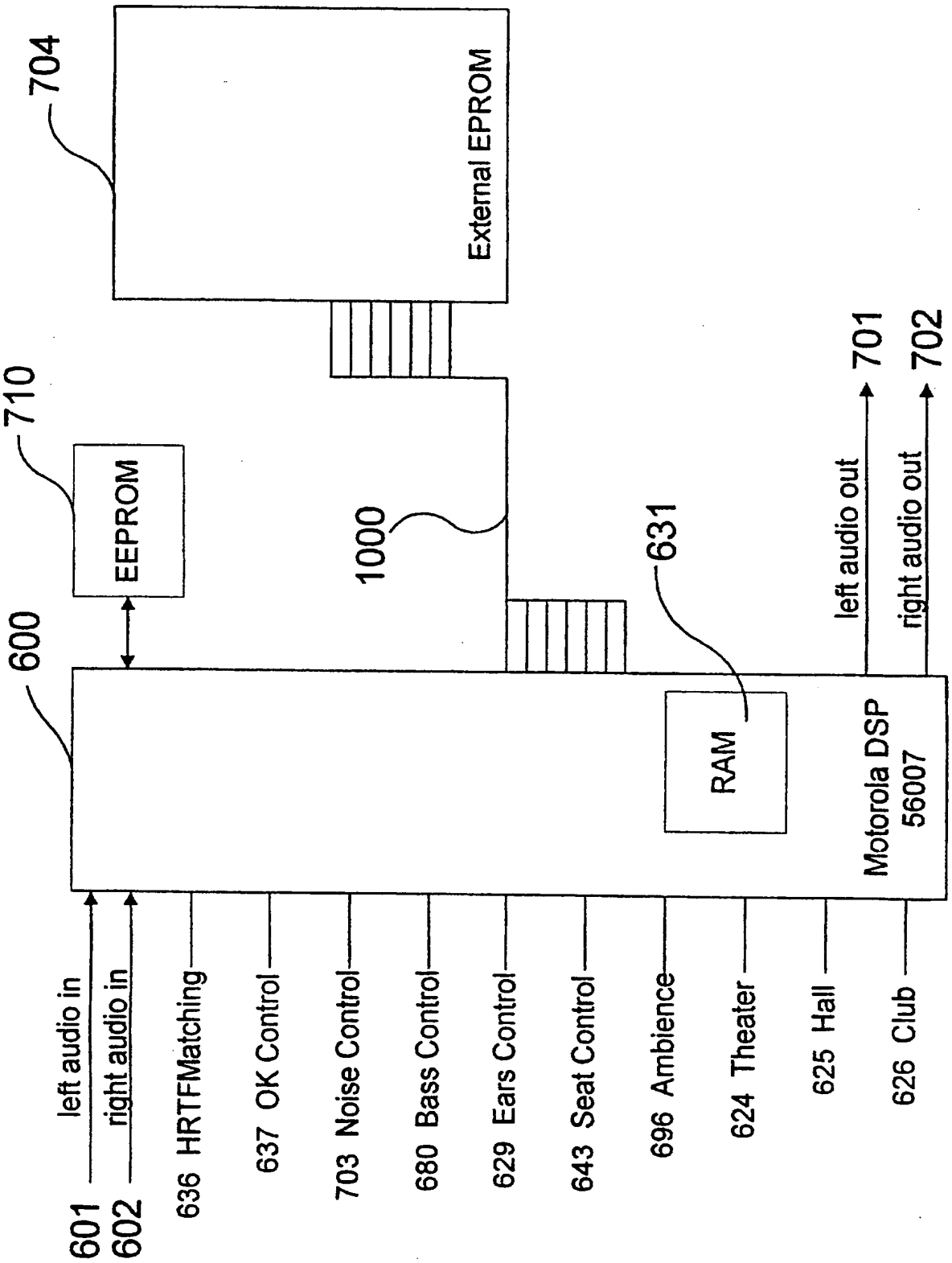
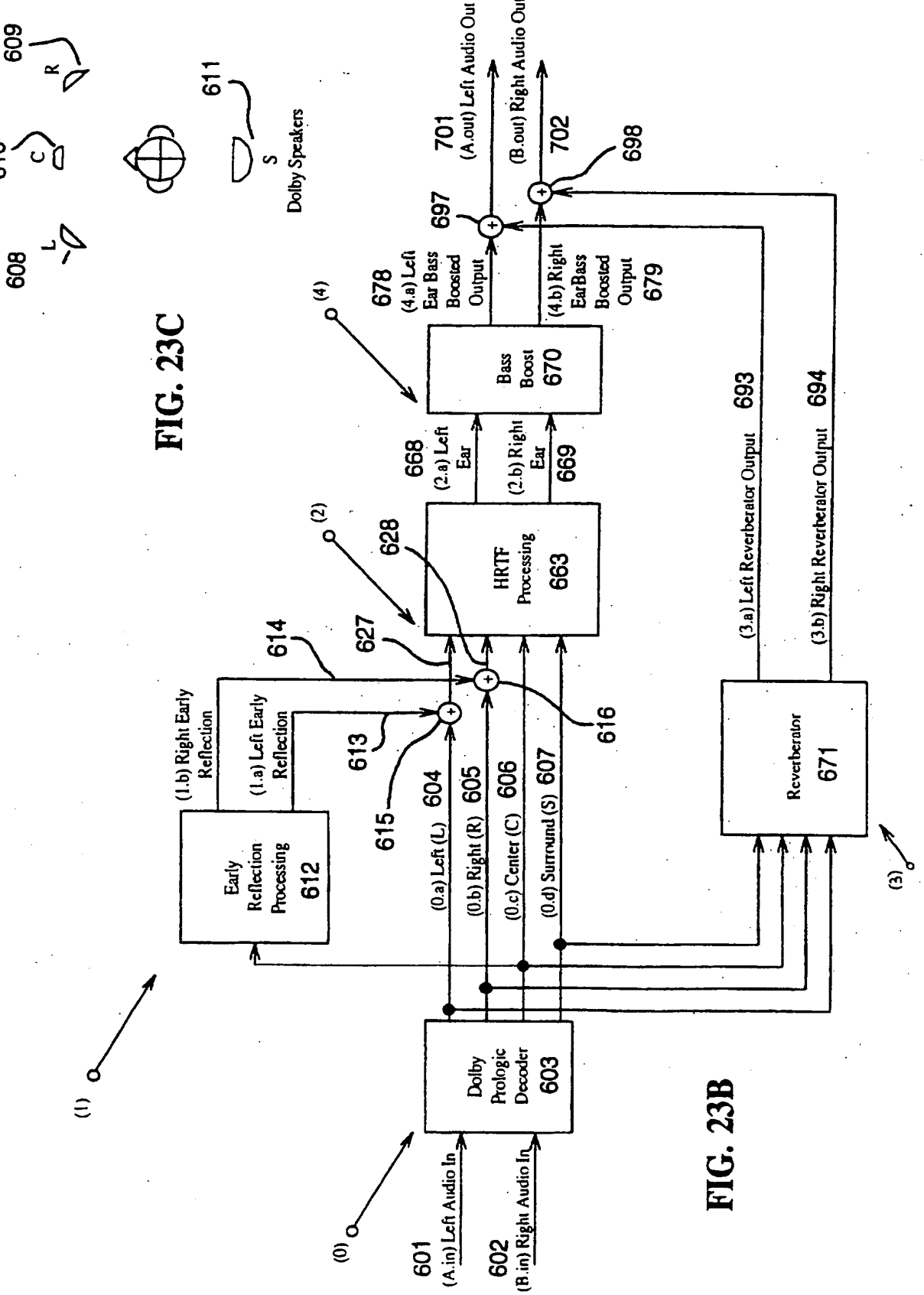

704
External EPROM
710
EEPROM
1000
631
600
RAM
Motorola DSP
56007
701
left audio out
702
right audio out
601
left audio in
602
right audio in
636 HRTFMatching
637 OK Control
703 Noise Control
680 Bass Control
629 Ears Control
643 Seat Control
696 Ambience
624 Theater
625 Hall
626 Club

FIG. 23A



3NSDOCID: <WO_____9725834A2_I_>

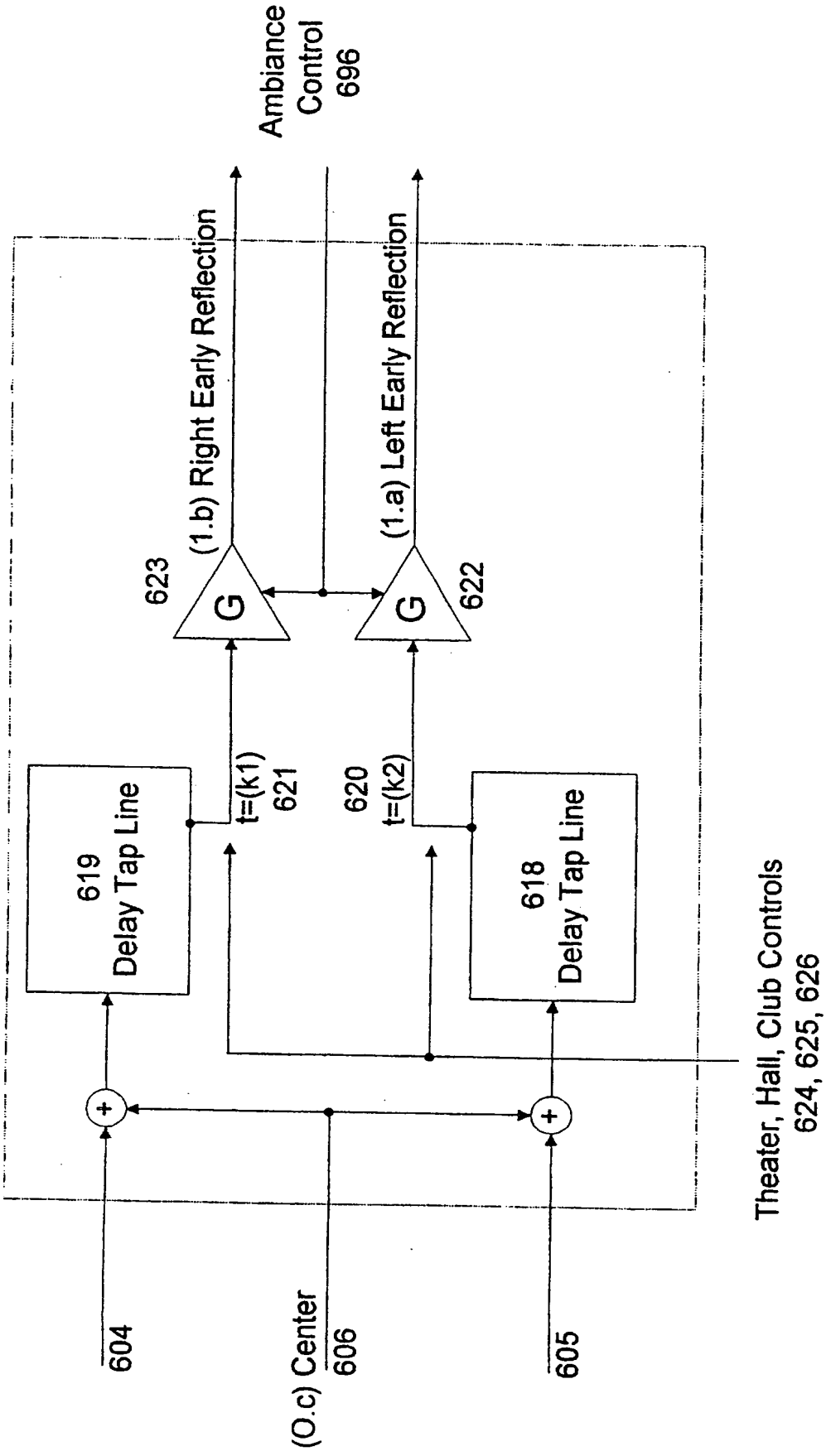
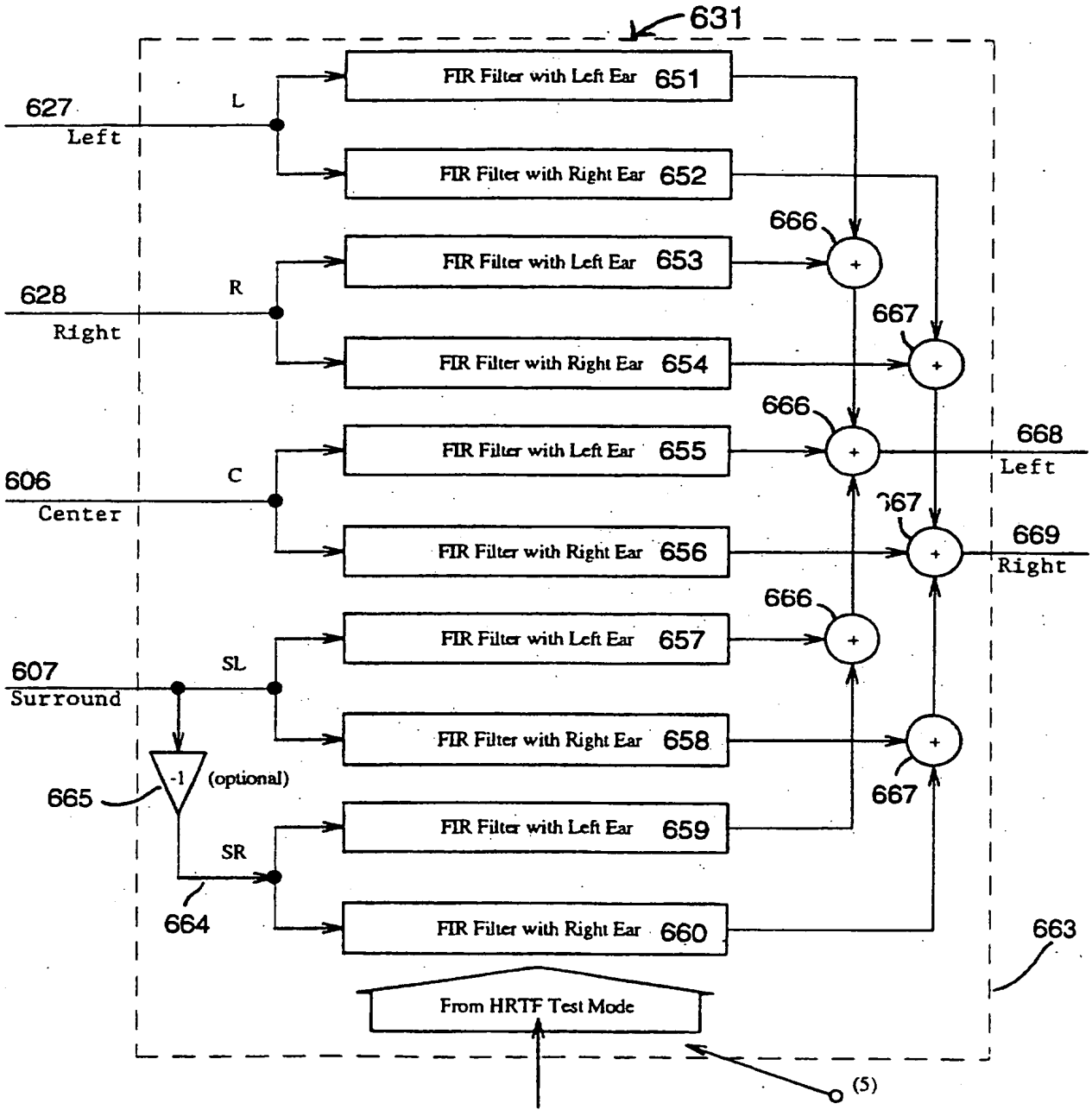


FIG. 24

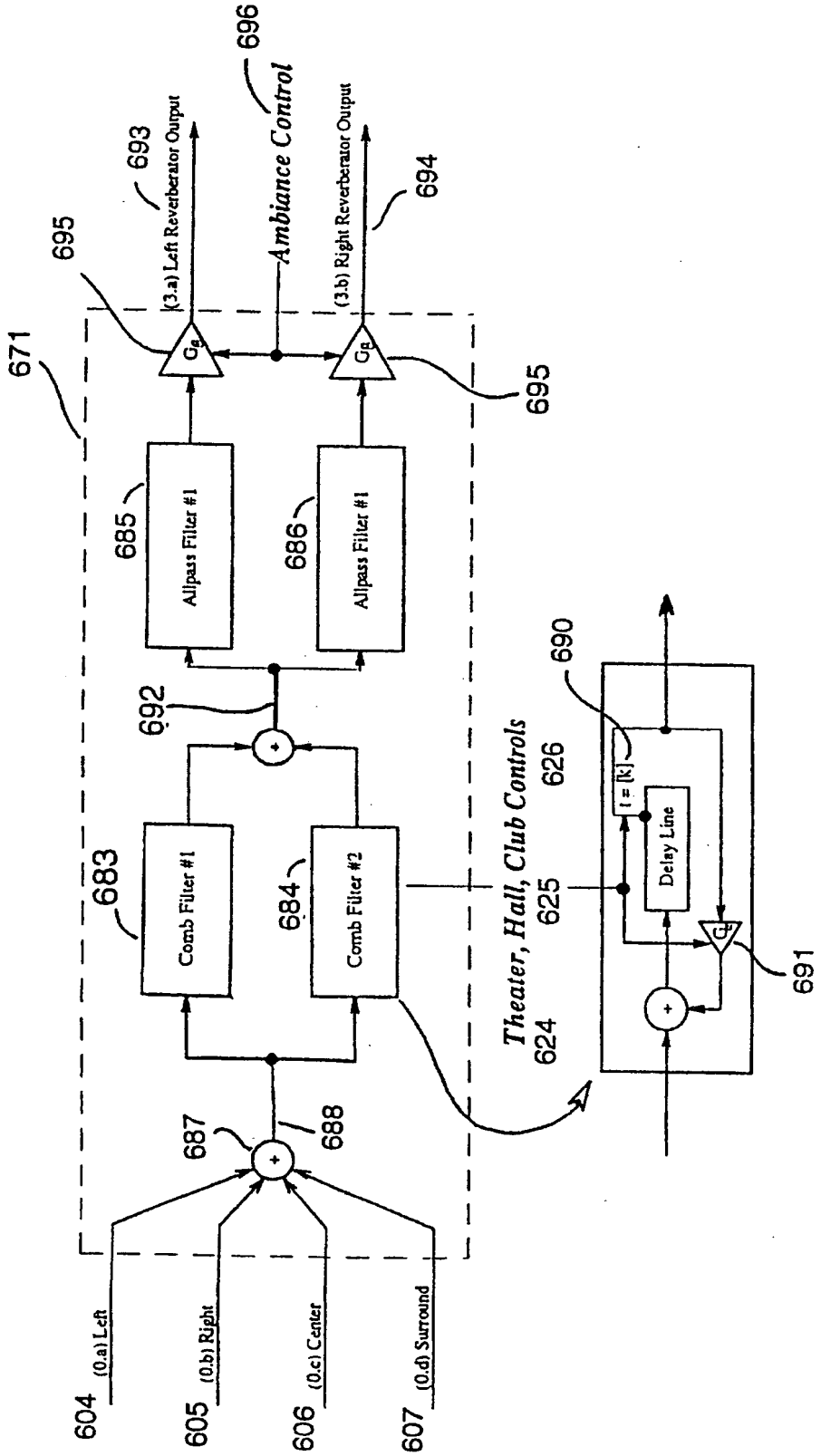
SDOCID: <WO_____9725834A2_I_>

FIG. 25



NSDOCID: <WO_____9725834A2_I_>

FIG. 26



ISDOCID: <WO_____9725834A2_I_>

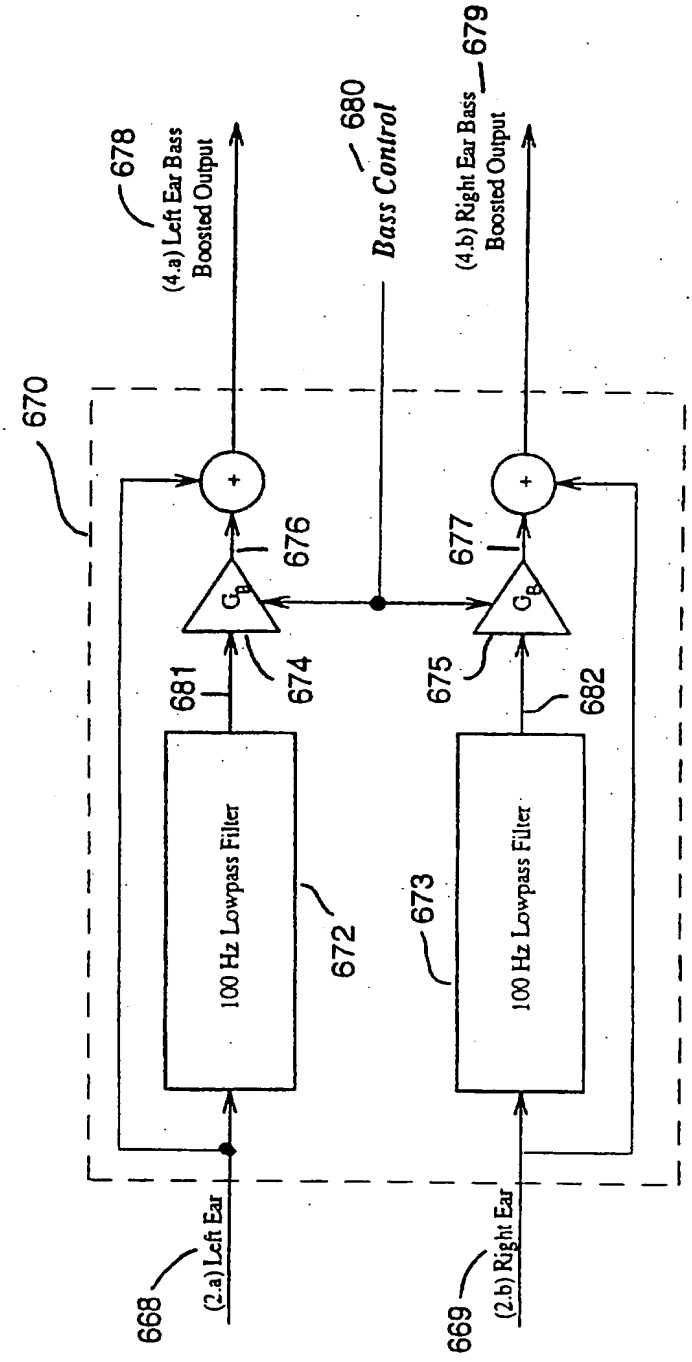


FIG. 27

NSDOCID: <WO_____9725834A2_I_>

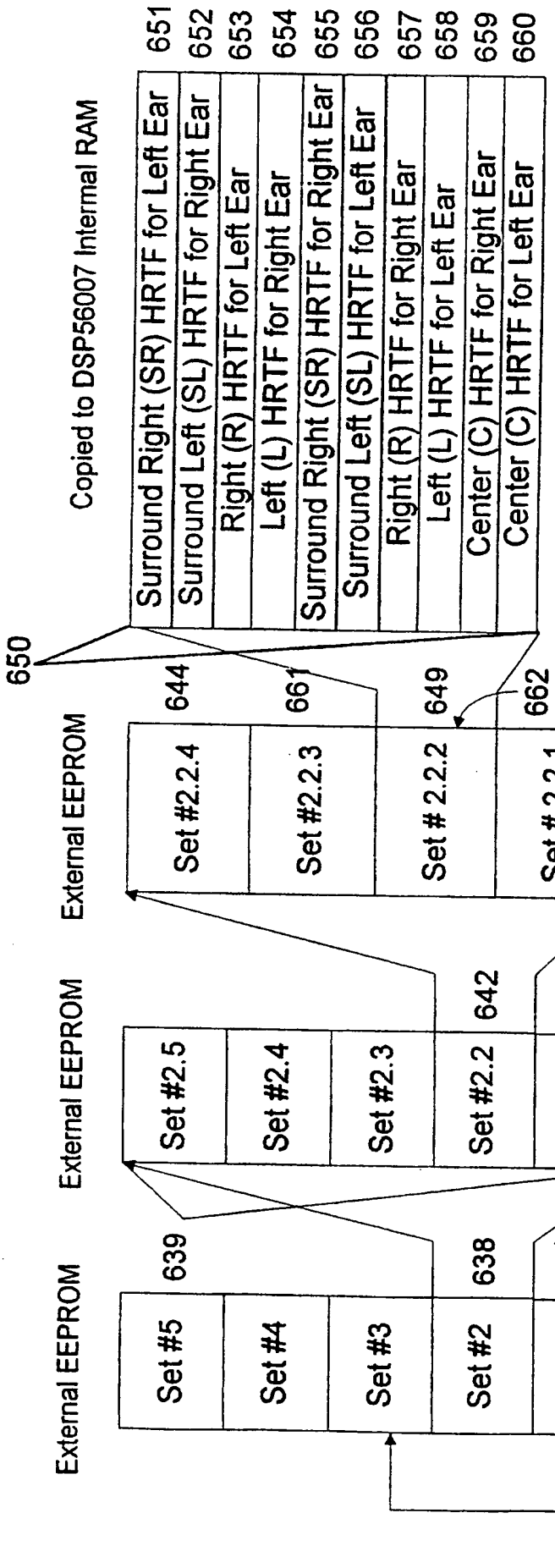

650
Copied to DSP56007 Internal RAM
651 Surround Right (SR) HRTF for Left Ear
652 Surround Left (SL) HRTF for Right Ear
653 Right (R) HRTF for Left Ear
654 Left (L) HRTF for Right Ear
655 Surround Right (SR) HRTF for Right Ear
656 Surround Left (SL) HRTF for Left Ear
657 Right (R) HRTF for Right Ear
658 Left (L) HRTF for Left Ear
659 Center (C) HRTF for Right Ear
660 Center (C) HRTF for Left Ear
External EEPROM
644 Set #2.2.4
661 Set #2.2.3
649 Set # 2.2.2
662 Set # 2.2.1
647
Seat Adjustment
Seat Control 643
External EEPROM
Set #2.5
Set #2.4
Set #2.3
642 Set #2.2
Set #2.1
HRTF Match Mode 2
OK Control 637
External EEPROM
639 Set #5
Set #4
Set #3
638 Set #2
Set #1
HRTF Match Mode 1 630
Ears Control 629
HRTF Matching Control 636

FIG. 28A

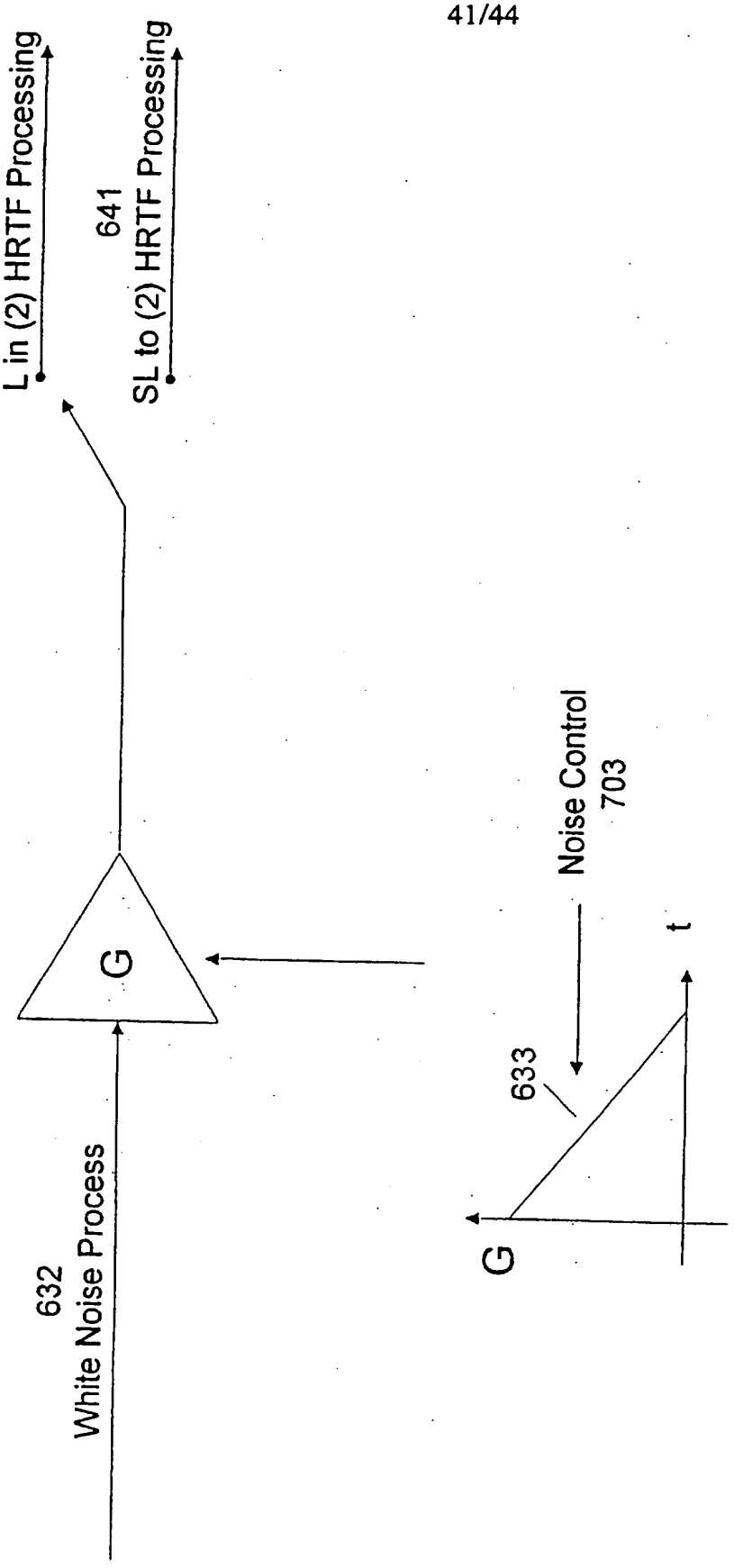

640
L in (2) HRTF Processing
641
SL to (2) HRTF Processing
G
632
White Noise Process
Noise Control
703
633
G
t

FIG. 28B

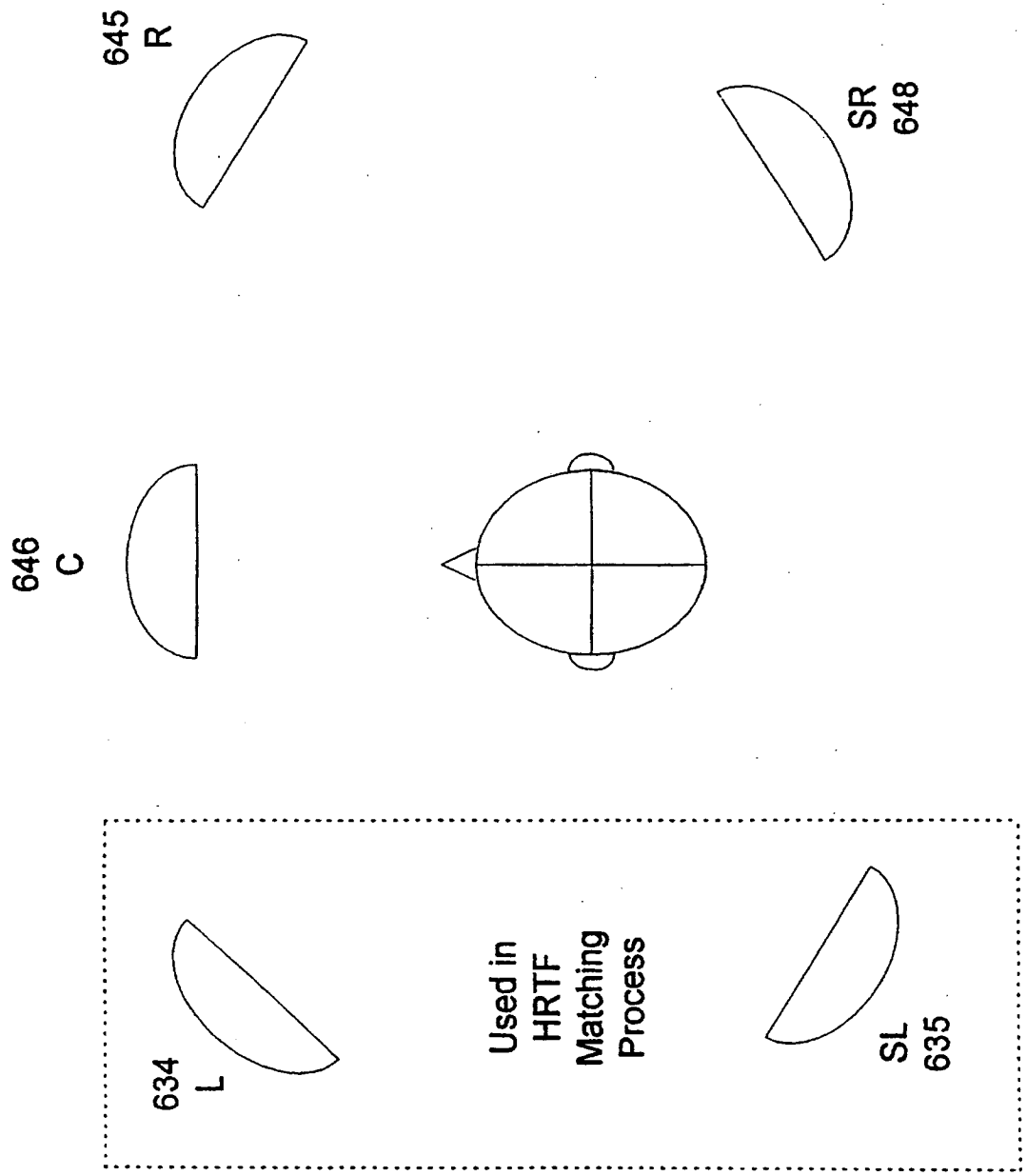


FIG. 28C

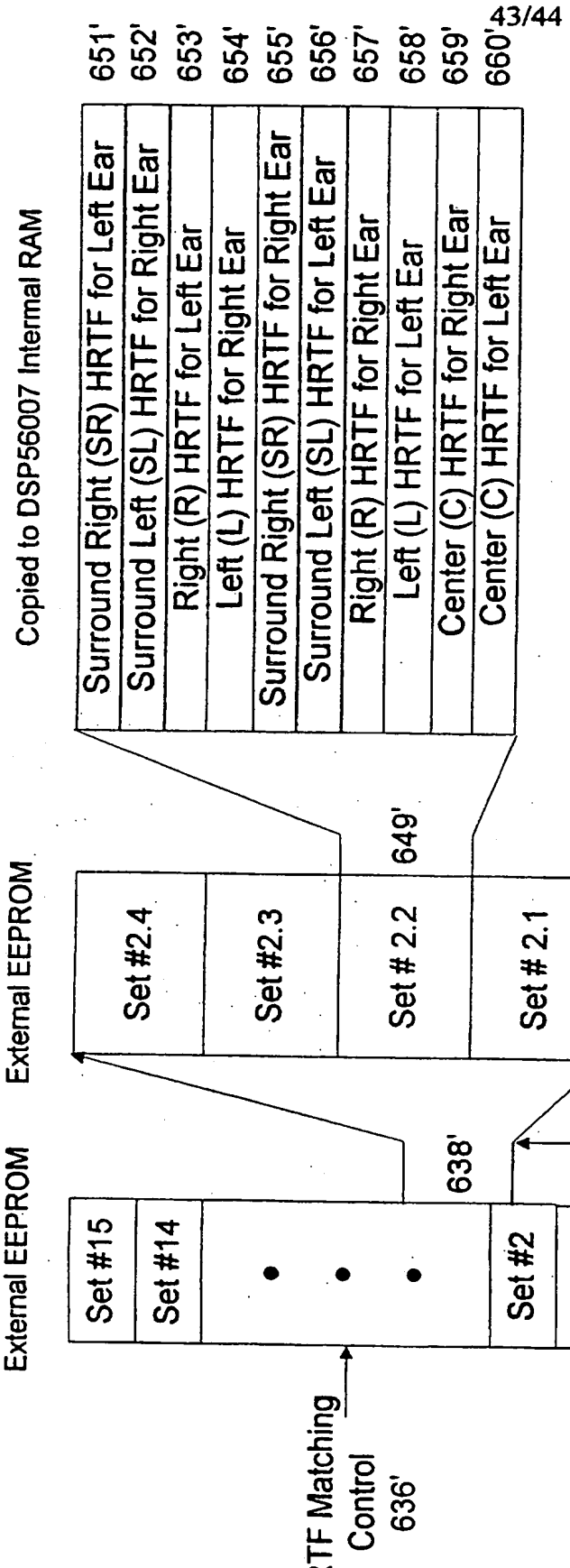


FIG. 29A

NSDOCID: <WO_____9725834A2_I_>

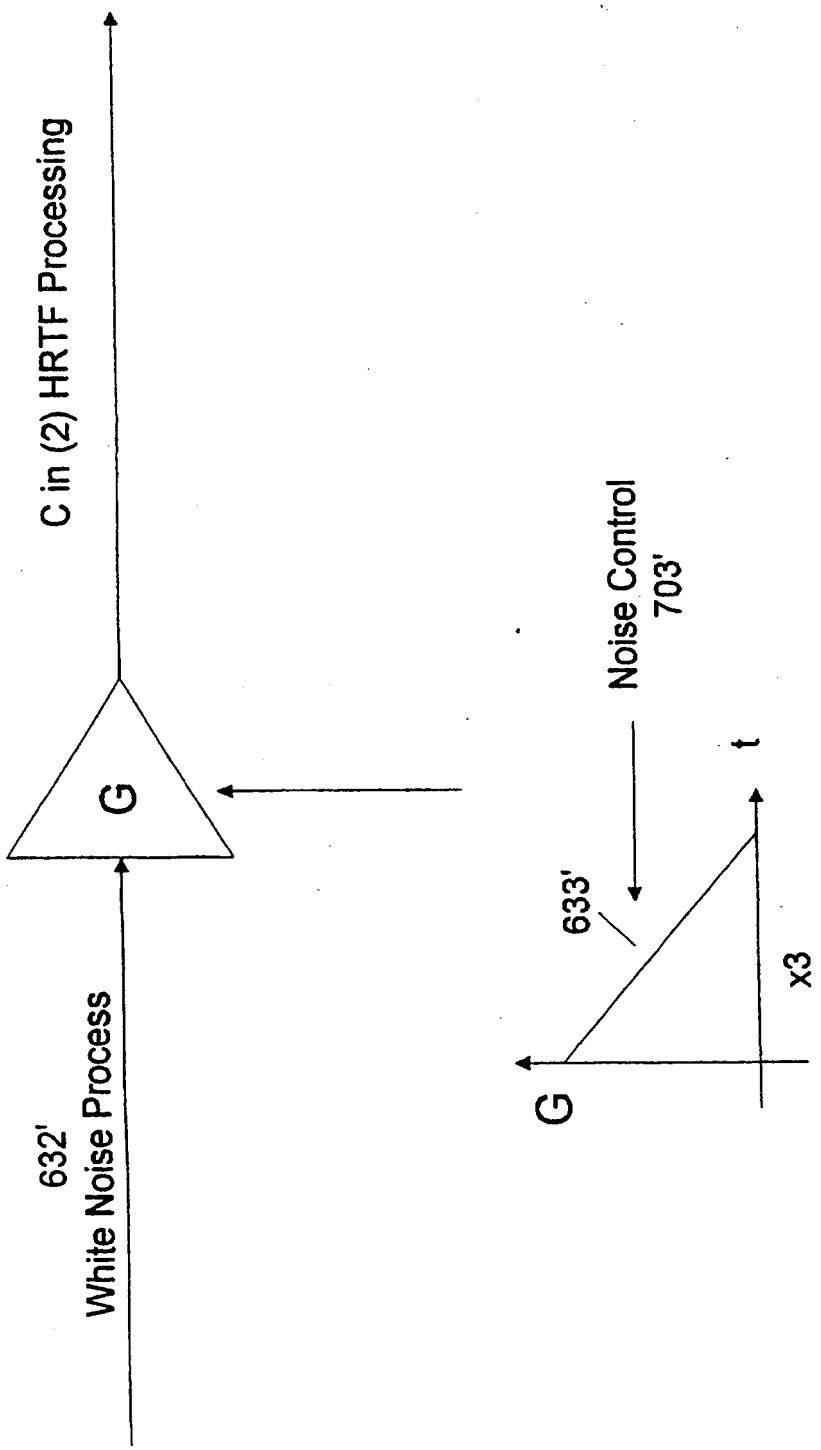

C in (2) HRTF Processing
G
632'
White Noise Process
Noise Control
703'
633'
G
t
x3

FIG. 29B

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04S 1/00, 5/00, 3/00 — A3

(11) International Publication Number: WO 97/25834

(43) International Publication Date: 17 July 1997 (17.07.97)

(21) International Application Number: PCT/US97/00145

(22) International Filing Date: 3 January 1997 (03.01.97)

(30) Priority Data:

08/582,830	4 January 1996 (04.01.96)	US
08/726,518	7 October 1996 (07.10.96)	US
08/726,590	7 October 1996 (07.10.96)	US

(71) Applicant *(for all designated States except US)*: VIRTUAL LISTENING SYSTEMS, INC. [US/US]; 4637 N.W. 6th Street, Gainesville, FL 32609 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: TUCKER, Timothy, J. [US/US]; 2215 N.W. 9th Place, Gainesville, FL 32609 (US). GREEN, David, M. [US/US]; 399 Federal Point Road, East Palatka, FL 32131 (US).

(74) Agents: PARKER, James, S. et al.; Saliwanchik, Lloyd & Saliwanchik, Suite A-1, 2421 N.W. 41st Street, Gainesville, FL 32606-6669 (US).

(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

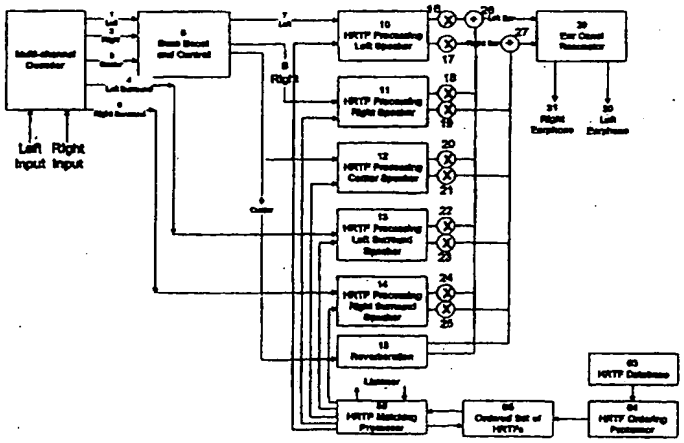
Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 18 September 1997 (18.09.97)

(54) Title: METHOD AND DEVICE FOR PROCESSING A MULTI-CHANNEL SIGNAL FOR USE WITH A HEADPHONE



(57) Abstract

A method and device processes multi-channel audio signals, each channel corresponding to a loudspeaker placed in a particular location in a room, in such a way as to create, over headphones, the sensation of multiple "phantom" loudspeakers placed throughout the room. Head Related Transfer Functions (HRTFs) are chosen according to the elevation and azimuth of each intended loudspeaker relative to the listener, each channel being filtered with an HRTF such that when combined into left and right channels and played over headphones, the listener senses that the sound is actually produced by phantom loudspeakers placed throughout the "virtual" room. A database collection of sets of HRTF coefficients from numerous individuals and subsequent matching of the best HRTF set to the individual listener provides the listener with listening sensations similar to that which the listener, as an individual, would experience when listening to multiple loudspeakers placed throughout the room. An appropriate transfer function applied to the right and left channel output allows the sensation of open-ear listening to be experienced through closed-ear headphones.

INSDOCID: <WO_____9725834A3_I_>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

ISDOCID: <WO_____9725834A3_I_>

INTERNATIONAL SEARCH REPORT

Intern al Application No
PCT/US 97/00145

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04S1/00 H04S5/00 H04S3/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04S H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 438 623 A (BEGAULT) 1 August 1995	1-7,15, 17,21, 32,38
A	see column 5, line 63 - column 7, line 27	22-26

X	EP 0 505 949 A (NIPPON TELEGRAPH & TELEPHONE) 30 September 1992	30
A	see page 17, line 24 - page 18, line 41	1,15

A	WO 95 23493 A (MOELLER ET AL.) 31 August 1995	1,15,30
	see page 5, line 2 - page 8, line 24	

A	WO 95 17799 A (CENTRAL RESEARCH LAB LTD ET AL.) 29 June 1995	1,15,30
	see page 2, line 17 - page 4, line 2	

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
4 July 1997	18.07.1997

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Gastaldi, G

Form PCT/ISA/210 (second sheet) (July 1992)

NSDOCID: <WO_____9725834A3_I_>

INTERNATIONAL SEARCH REPORT

Interna al Application No
PCT/US 97/00145

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4 739 513 A (KUNUGI ET AL.) 19 April 1988 see column 5, line 10 - column 8, line 26; figures	1,15
A	WO 95 31881 A (CRYSTAL RIVER ENGINEERING INC) 23 November 1995 see page 8, line 20 - page 15, line 2	1,15,30
A	US 5 404 406 A (FUCHIGAMI ET AL.) 4 April 1995 see column 4, line 20 - column 6, line 7	1,15,30
A	US 5 371 799 A (LOWE ET AL.) 6 December 1994 see column 2, line 2 - line 48	1,15,30

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

NSDOCID: <WO_____9725834A3_I_>

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern al Application No
PCT/US 97/00145

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5438623 A	01-08-95	NONE	
EP 0505949 A	30-09-92	JP 4295728 A	20-10-92
		JP 7039968 B	01-05-95
		DE 69207039 D	08-02-96
		DE 69207039 T	12-09-96
		US 5187692 A	16-02-93
WO 9523493 A	31-08-95	AU 1755595 A	11-09-95
		CA 2184160 A	31-08-95
		EP 0746960 A	11-12-96
WO 9517799 A	29-06-95	EP 0737410 A	16-10-96
US 4739513 A	19-04-88	JP 60254997 A	16-12-85
		JP 1797135 C	28-10-93
		JP 5003928 B	18-01-93
		JP 61108214 A	26-05-86
		JP 7054998 B	07-06-95
		JP 61108294 A	26-05-86
		EP 0165733 A	27-12-85
WO 9531881 A	23-11-95	AU 2460395 A	05-12-95
		EP 0760197 A	05-03-97
US 5404406 A	04-04-95	JP 6178397 A	24-06-94
		JP 6178398 A	24-06-94
US 5371799 A	06-12-94	NONE	

ISDOCID: <WO_____9725834A3_I_>

THIS PAGE BLANK (USPTO)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☑ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)